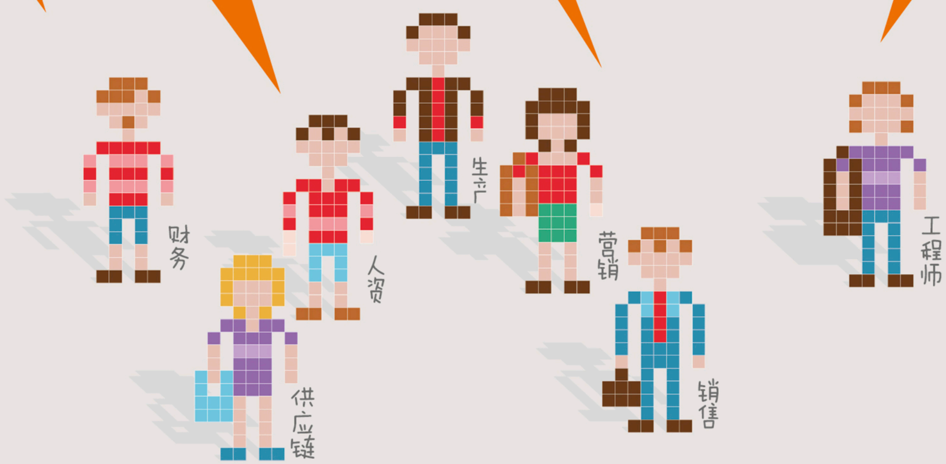
图灵原创

刘红阁 王淑娟 温融冰◎著

人人都是数据分析师

Tableau应用实战



拖 放 成 图　　探寻统计规律　　便捷数据管理

内置最佳做法　　海量数据读取　　交 互 式 地 图

中国工信出版集团　人民邮电出版社
POSTS & TELECOM PRESS

数字版权声明

图灵社区的电子书没有采用专有客户端，您可以在任意设备上，用自己喜欢的浏览器和PDF阅读器进行阅读。

但您购买的电子书仅供您个人使用，未经授权，不得进行传播。

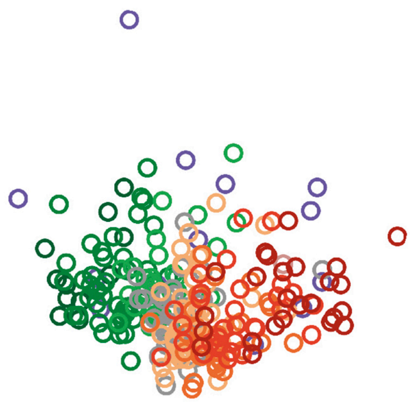
我们愿意相信读者具有这样的良知和觉悟，与我们共同保护知识产权。

如果购买者有侵权行为，我们可能对该用户实施包括但不限于关闭该帐号等维权措施，并可能追究法律责任。

刘红阁博士，百度凤巢分析经理，专注于数据挖掘、机器学习、数据可视化等领域。

王淑娟，埃森哲咨询顾问，专注于电力行业运营咨询、数据分析及数据可视化等领域。

温融冰，埃森哲咨询经理，专注于企业价值管理、预算与绩效管理及数据分析等领域。



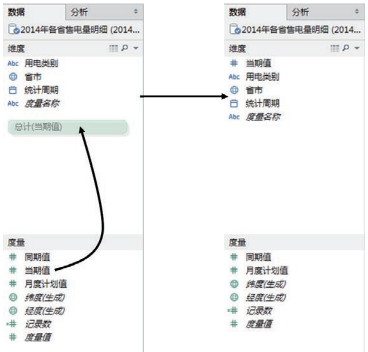
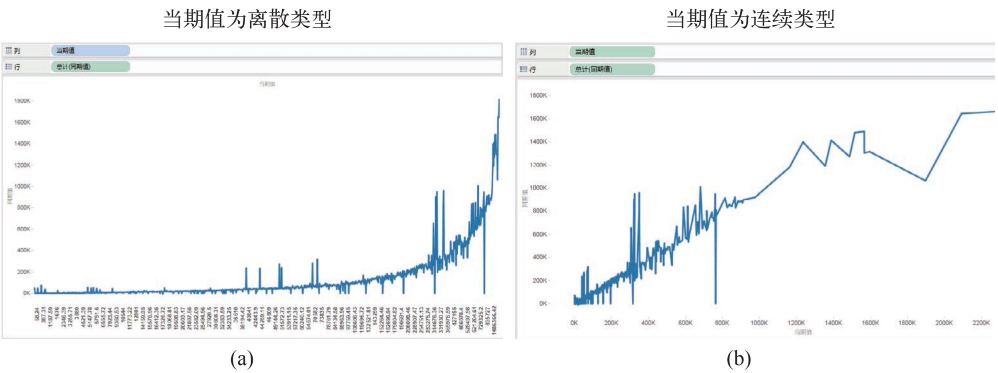


图 2-5　维度与度量转换



(a)　(b)

图 2-6　离散和连续类型

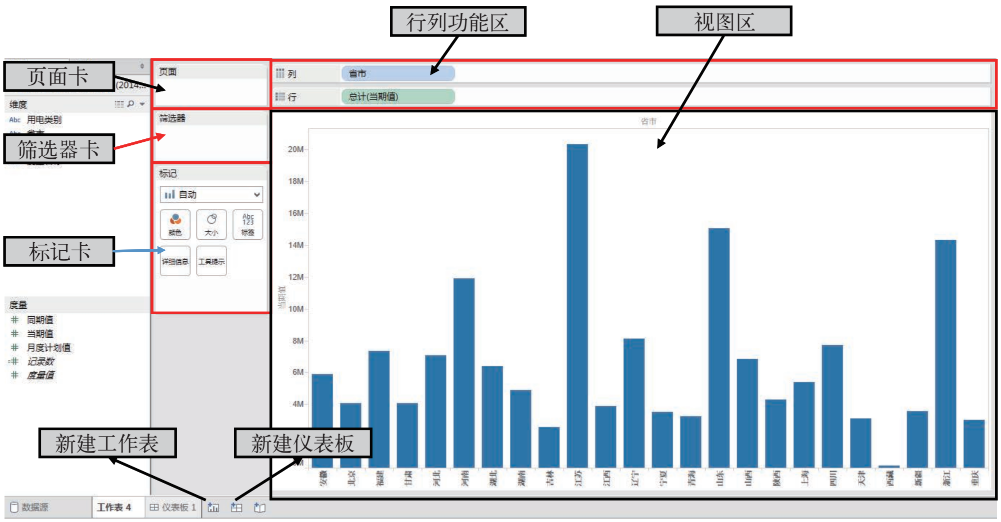


图 2-8　认识视图工作区

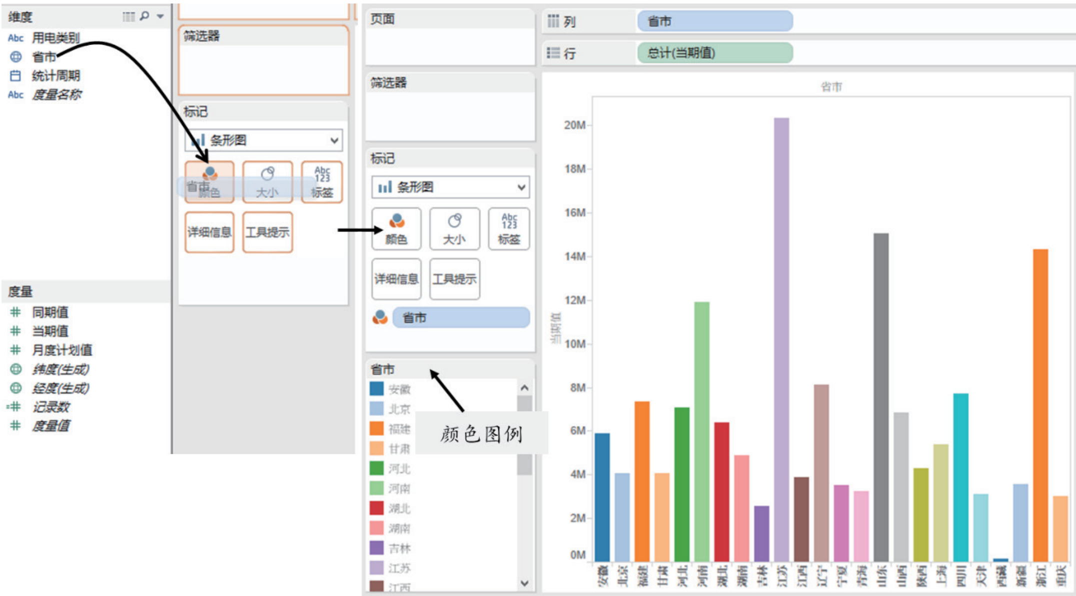


图 2-16　颜色图例

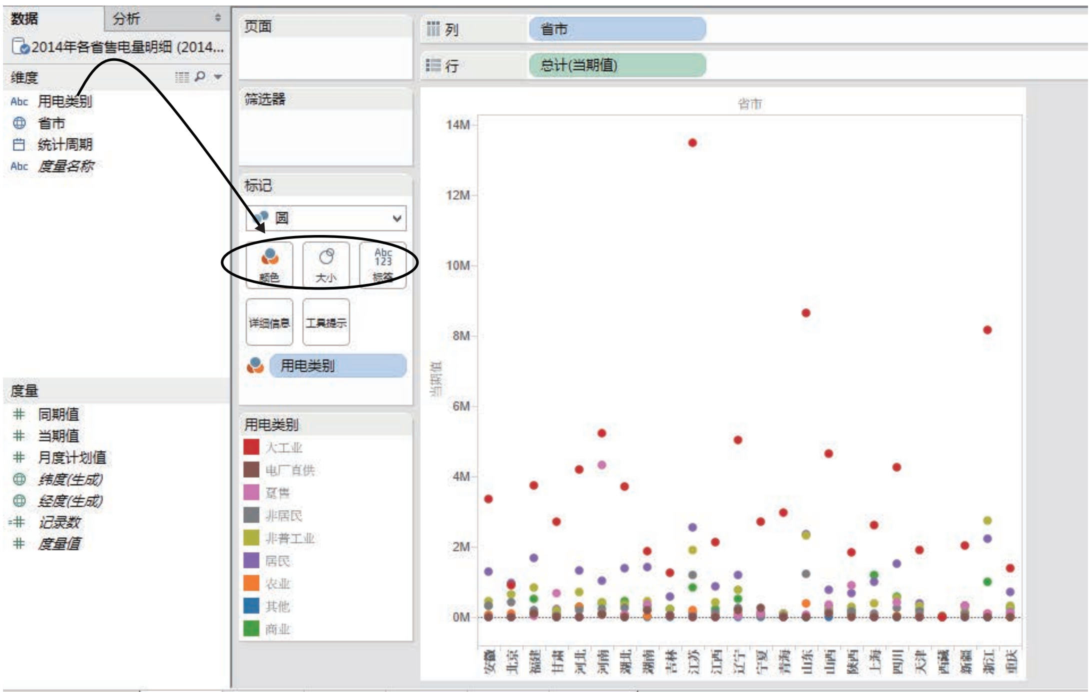


图 2-23　依据颜色的详细信息

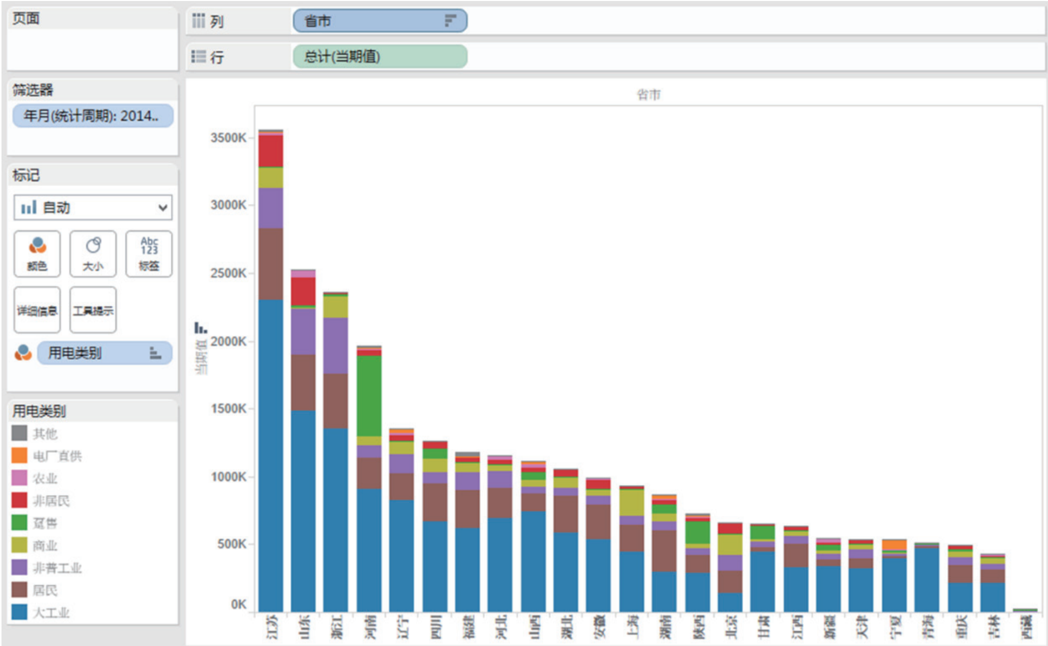


图 4–6　排序后的 2014 年 3 月各省市售电量用电类别分布图

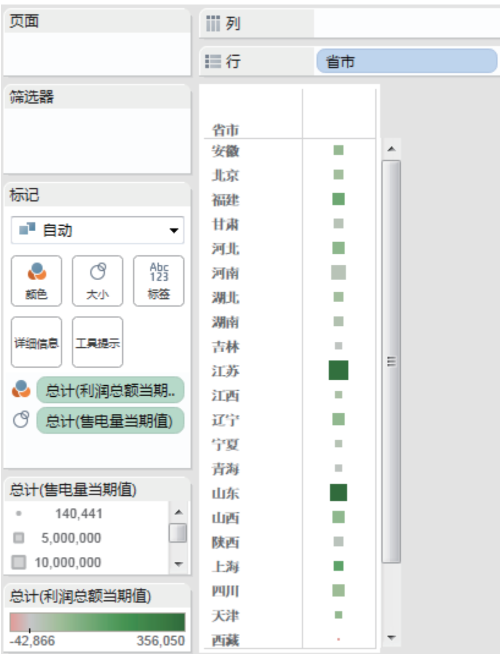


图 4-33　压力图——2014 上半年各地市累计售电量与利润总额情况



图 4-34　一个指标的突显表——2014 年上半年各省市累计利润总额情况表

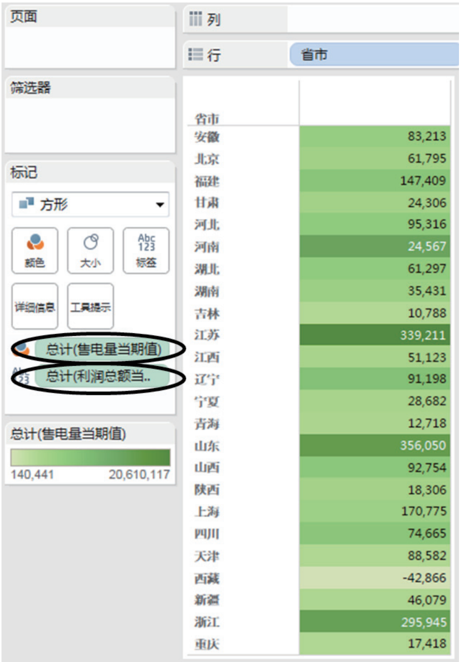


图 4-35　两个指标的突显表——2014 年上半年各省市累计售电量与利润总额情况表

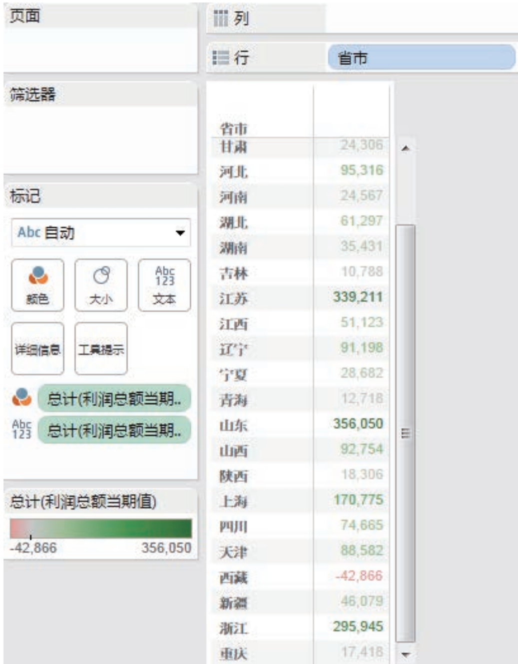


图 4-36　利润总额按数值大小用不同颜色显示

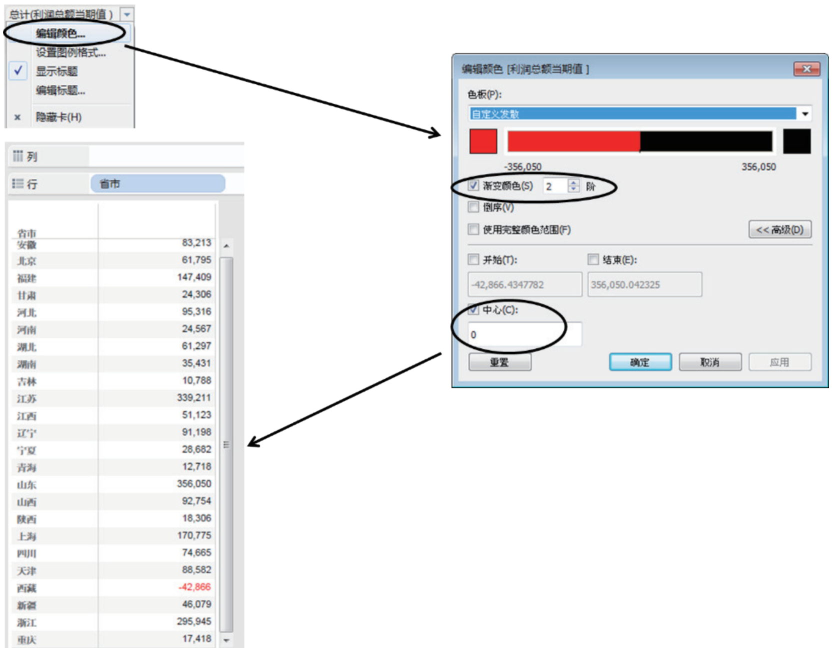


图 4-37　突显表——突出显示利润总额的负值

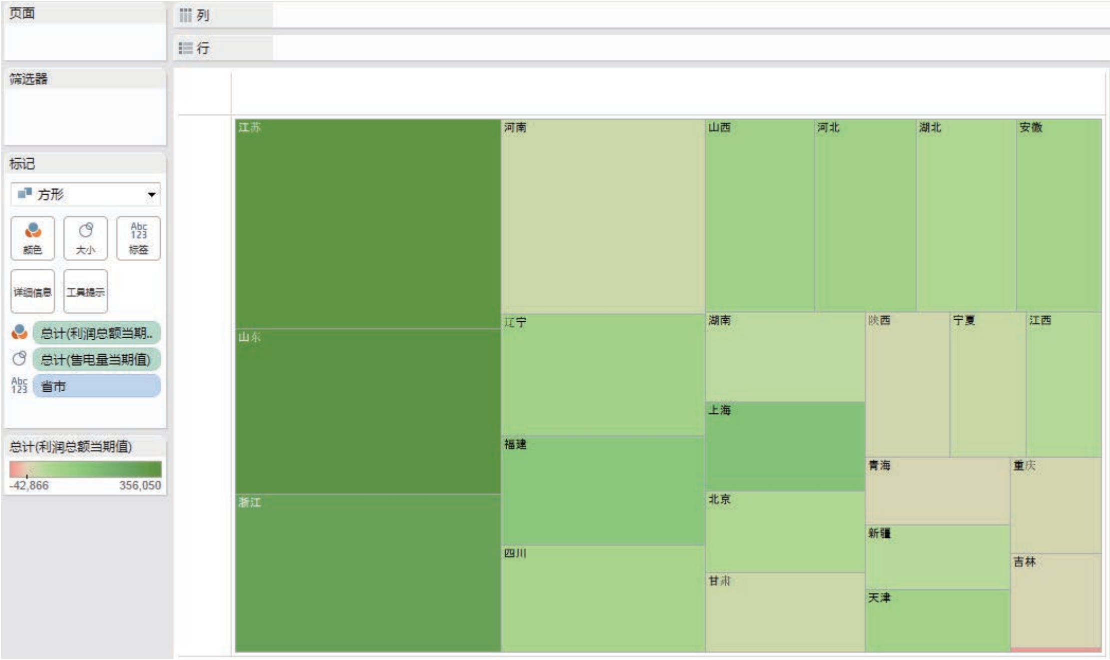


图 4-38　树地图——2014 年上半年全国各省市累计售电量与利润总额关系

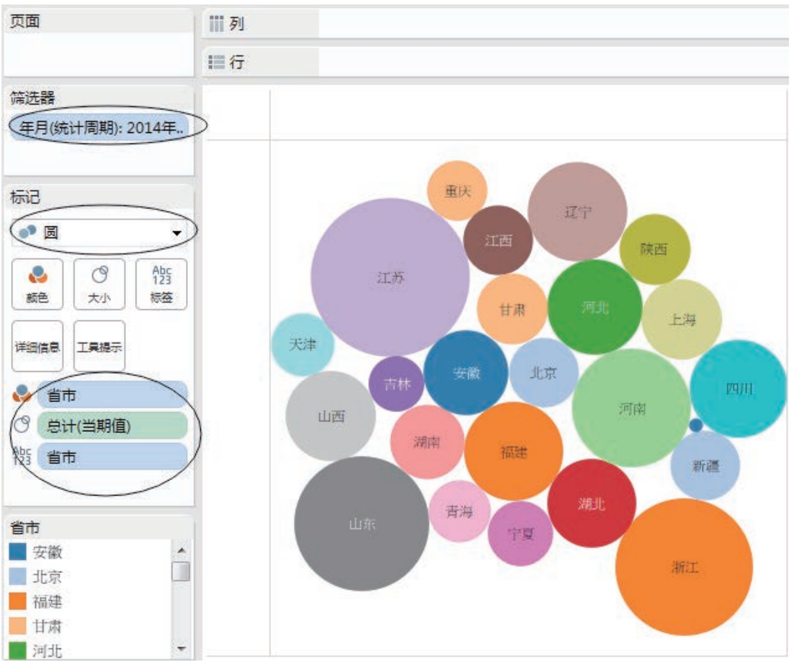


图 4-39　填充气泡图——2014 年 6 月各省市售电量情况

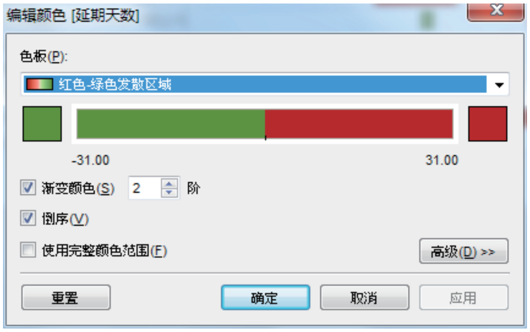


图 4-48　编辑“延期天数”的颜色

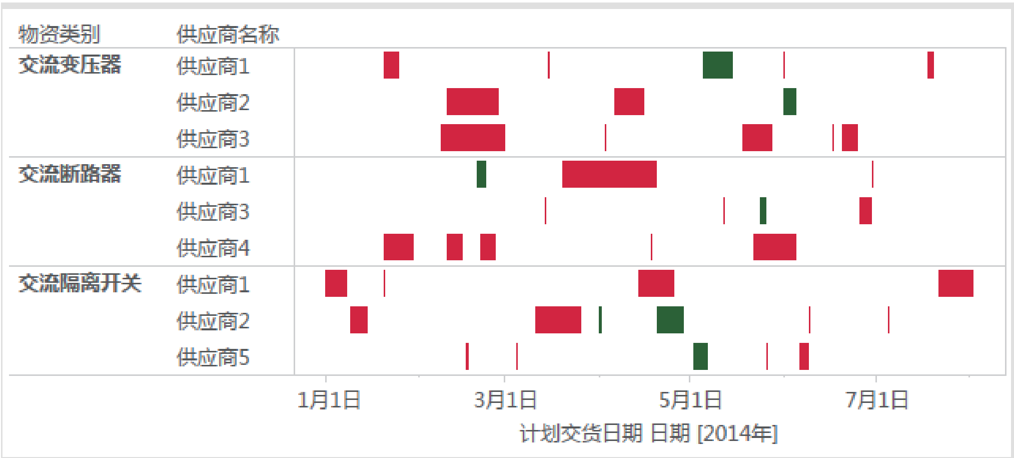


图 4-49　供应商及时供货情况分析

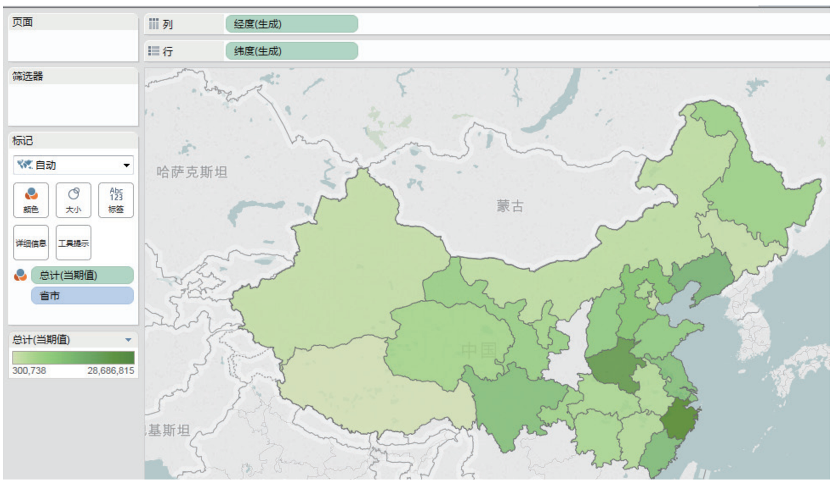


图 5-15　填充地图

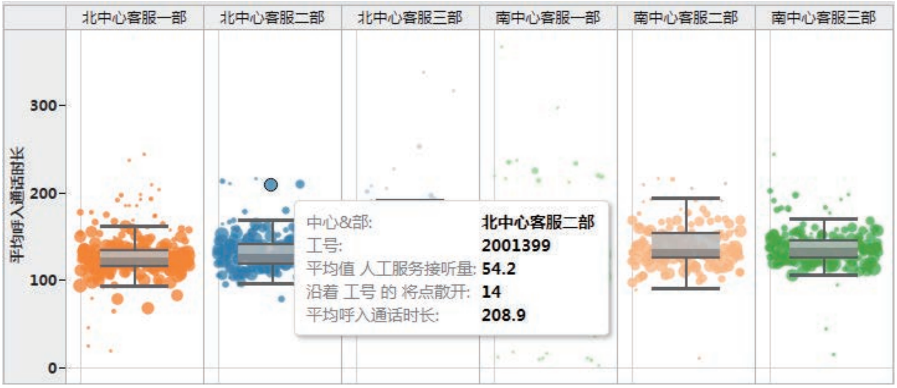


图 7-17 设置将点散开效果

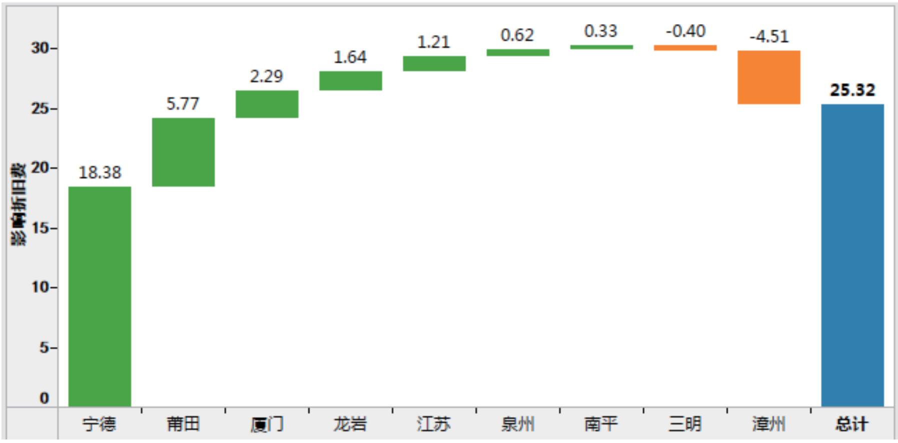


图 7-18 瀑布图

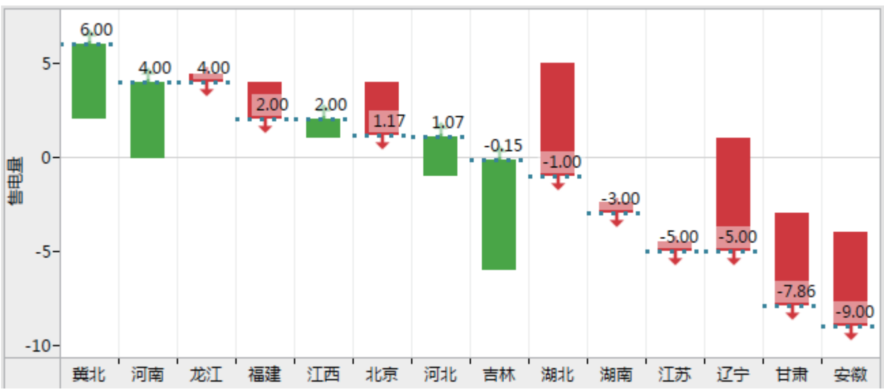


图 7-24 “变化排序”图

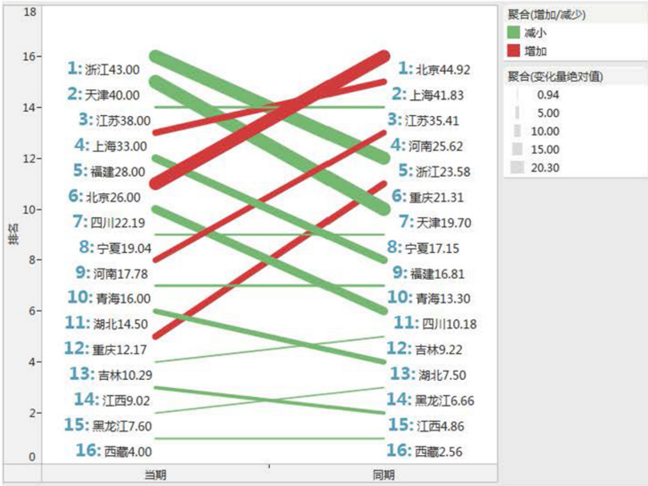


图 7-32　倾斜图

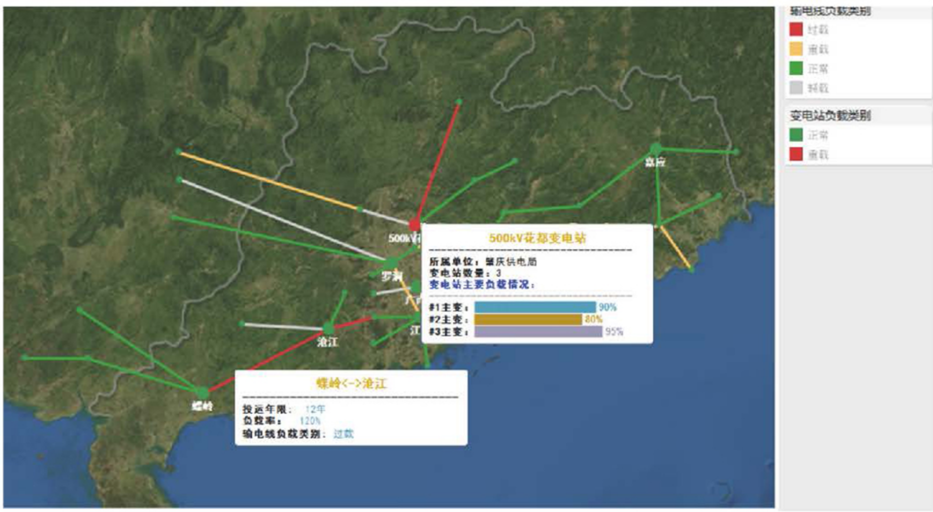


图 7–37　设置图形的颜色和形状

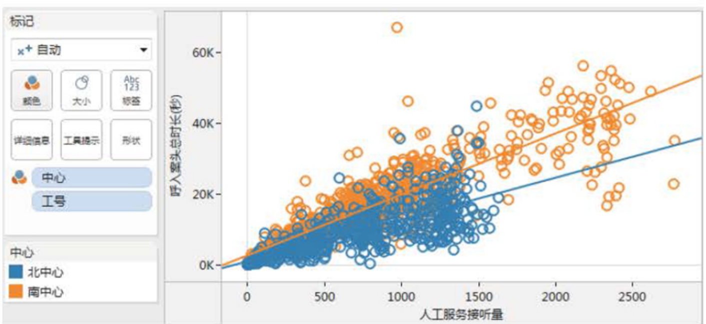


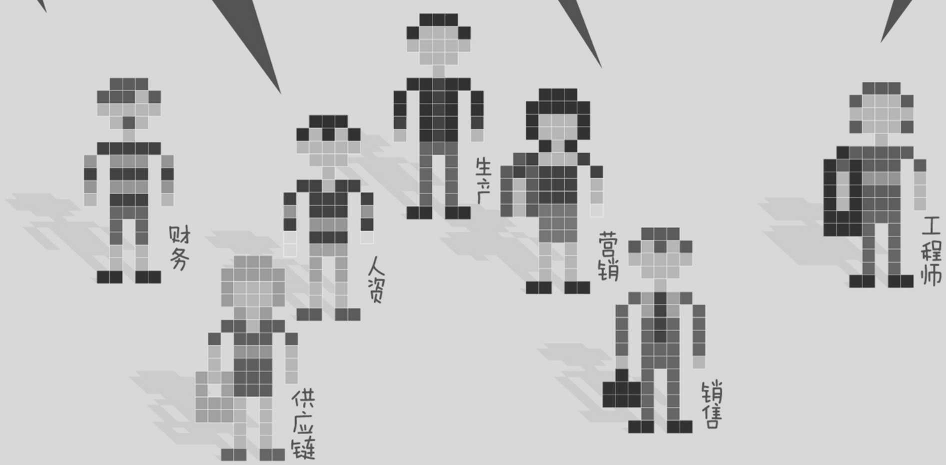
图 8-11　按颜色绘制趋势线

TURING 图灵原创

刘红阁 王淑娟 温融冰◎著

人人都是数据分析师

Tableau应用实战



人 民 邮 电 出 版 社

北 京

图书在版编目（CIP）数据

人人都是数据分析师 : Tableau应用实战 / 刘红阁,
王淑娟, 温融冰著. -- 北京 : 人民邮电出版社,
2015.11
（图灵原创）
ISBN 978-7-115-40686-6

Ⅰ. ①人… Ⅱ. ①刘… ②王… ③温… Ⅲ. ①可视化
软件 Ⅳ. ①TP31

中国版本图书馆CIP数据核字(2015)第241018号

内 容 提 要

本书基于 Tableau 9.1 最新版本编写，详细介绍了 Tableau 的数据连接与编辑、图形编辑与展示功能，包括数据连接与管理、基础与高级图形分析、地图分析、高级数据操作、基础统计分析、如何与 R 集成进行高级分析、分析图表整合以及分析成果共享等主要内容。同时，书中以目前电力行业已有的监测、分析业务实践为基础，以丰富的实际案例贯穿始终，对各类方法、技术进行了详细说明，方便读者快速掌握数据分析方法。

本书适用于互联网、银行证券、咨询审计、快消品、能源等行业数据分析用户以及媒体、网站等数据可视化用户。

◆ 著　　　　刘红阁　王淑娟　温融冰
责任编辑　王军花
策划编辑　张　霞
责任印制　杨林杰

◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路11号
邮编　100164　　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
北京　　　　　　印刷

◆ 开本：800×1000　1/16
印张：21.5　　　　彩插4
字数：502千字　　　　2015年 11 月第 1 版
印数：1－4 000册　　　　2015年 11 月北京第 1 次印刷

定价：69.00元

读者服务热线：(010)51095186转600　印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京崇工商广字第 0021 号

Tableau用户评价

“Tableau 真正地帮助我们快速发展和创造价值。对于生产，这种帮助无时无刻不在体现。”

——Will Bishop，特斯拉汽车公司高级测试工程师

“Tableau 帮助我们以更少的资源清理和整塑用于报告的数据，因为实现了报告自动化，所以我们这个小小的团队已经将大量资源集中在了将数据分析转变成见解上。”

——Secil Watson，富国银行集团执行副总裁

“我们完全钟情于 Tableau，钟情于它对提供组织中的每个人而不只是分析师和高级用户使用现成、可操作的数据的承诺。”

——David Baudrez，思科公司欧洲、中东和非洲地区商业洞察主管

“如果一次又一次地重复同一件事情，我永远也无法跳出固有的思维模式。Tableau 可以让我们以一种完全不同的方式来查看数据。”

——Kevin King，可口可乐装瓶公司报告与分析主管

“使用 Tableau 之后，去年我们的收入增加了 2 亿美元，增幅达到了 2%。Tableau 很实用，使用也很简单，而且比其他商业智能工具反应快。要想在航空业或者运输业赚到钱，就得用 Tableau。”

——James Pu，中国东方航空公司网络与营收部高级主管

“如果现在没有 Tableau 系统，我们可能仍在使用电子表格软件，那样就需要委派专人进行数据分析，不仅偶尔的人为错误不可避免，而且当分析员没有上班时，当天的报告可能也无法获得。”

——Leon Zhang，汉堡王（中国）信息技术部副总监

序一

古谚有云“一图胜万言”，这在不同文化中均广为人知。有人经考证认为，其最早来自记者弗莱德·巴纳德（Fred R. Barnard）1927年发表在一本贸易杂志上的文章；然而另一些人认为它源于中国的孔子，这位2000多年前的智者曾讲过“书意能达万言”。无论出处如何，专家们已经一次又一次地发布研究成果来证实这个结论：人类主要是视觉动物。当我写这篇序言时，我仍惊叹于这种认知在中国是多么地深入人心，从远古时代雄伟的绘画到象形文字、形声字，再到今天的简体字，都有着广泛的体现。

而Tableau有幸基于这样的智慧创造应用产品，来帮助更多的人读懂他们的数据。我们十多年前起源于斯坦福大学的这个愿望，已经惠及150多个国家3万多家机构及成千上万的个人。当今世界，人类前所未有地连接在一起，而未来5年内，将会产生多如繁星的数据，那么能够理解、利用和掌握自身的数据将变成一项最基本的生存技能。

许多机构已经开始面对这个人人自助分析、数据驱动决策的现实。少数比较成功，但更多疲于追赶。Tableau 亚太区的一项市场调研发现，那些能够满足业务人员自助开展数据读取和分析需求的企业，其业绩比采取传统做法的企业好两倍。但是尽管知晓此好处，75%的业务人员表明他们自助开展数据分析的需求仍未充分满足或完全未满足。

中国经济的发展早已超越曾经的“四小龙”，在知识经济与数字化时代，越来越多的国有企业、私营企业甚至公共服务机构都面临同样的挑战。培训员工应用自助式的分析工具发现海量数据中的价值已变得势在必行。

而这正是本书的关键所在！尽管市面上已经有了传统的商业智能解决方案，但是面对海量的数据和整合这些解决方案的复杂性，很多中国企业仍疲于应对。本书通过真实的案例，阐述了一个完全不同于以往的数据分析方法论。它展示了领先企业如何让商业智能不再局限于少数技术人员，让多数人都掌握自助分析，读懂数据，创造更大的价值。

中国是世界上最大的能源消费国，电力行业的高效运营有着重大而广泛的影响。我相信读者会从电力行业敏捷数据分析的最佳实践中受益匪浅。关于这一点，我非常感谢我们的合作伙伴埃森哲的三位作者，他们持续为客户带来行业洞见，并先见地倡导客户应用Tableau来满足数据分析的需求。

JY Pook

Tableau 亚太区副总裁

序二

随着数字化技术的广泛应用和信息化水平的持续提升，我们正在逐步迈入大数据时代。全球发达国家已经充分认识到这一发展趋势，纷纷将大数据上升为国家战略，2015 年 8 月我国国务院通过了《关于促进大数据发展的行动纲要》，为其发展推波助澜。在企业层面，数据已成为一种新型的重要战略资产，越来越多的企业开始强化数据资产管理，建立专门的组织或明确相关的责任部门，大力拓展数据分析与应用。数据分析已经成为一门显学，也正在发展成为一项独立的业务，发挥着越来越大的作用。

在埃森哲为客户提供的多种咨询服务中，数据分析咨询数量自 10 年前开始逐渐变多。在国内，随着诸多企业信息系统建设的不断发展，近几年利用数据、信息进行业务监测管控与管理优化的需求日益显现。然而由于对数据分析这一新鲜事物认识不足，经常出现这样的情况：尽管在服务器、数据库、分析软件与技术工具的投资颇多，但却难以取得实用效果。最为明显的认识偏差是没有将数据分析视为一项业务能力，有的人把数据分析等同于报表、报告等结果性的东西，另外有些人又把它视为一些分析技术、工具与 BI 系统。实质上，数据分析有自己独特的策略、组织、流程、技术组成与要求，并且各要素之间需要相互支撑与配合，才能持续产生满足需要的信息与见解。不止如此，数据分析作为一项业务，具有明显不同于一般流程性业务的显著特征：频繁变化。随着企业战略发展、管理焦点与数据支撑度的变化调整，数据分析与使用需求也要不断变化才能满足需要，传统流程性业务的很多做法在数据分析业务上就不能一成不变。一言以蔽之，数据分析是一种随需而变的新型业务。

数据分析业务变化多，频度高，处理数据量日益庞大，这给企业提出了严峻挑战，尤其是在数据处理与分析支撑技术层面。我们很多客户在这个方面决策有误，采用了不那么有效的技术路线，引入或建设了对数据分析工作支撑度有限的技术与系统，最终限制了数据分析能力与作用结果。在给客户提供数据分析咨询的过程中，埃森哲先后使用过 SAS、SPSS、BI、BO、Excel 等多种复杂程度各异的工具与系统。基于我们的使用体会与客户反馈，从以最小化的投入来最大化满足当前阶段企业数据分析需求的角度看，Tableau 不失为一个值得推荐的选择。

Tableau 既可以作为一个加强版的类 Excel 工具，供数据分析人员对数据进行交互式、可视化分析和挖掘，提高分析工作效率，给出有意义的分析结论；也可以将有意义的分析路径、成果进行发布，共享给其他关联或管理人员进行浏览、查询与使用，实现分析人员与业务人员的协同。在数据分析领域，Tableau 在国外领先公司中得到了广泛应用，并且在国内的市场规模也日益壮大。当然，数据分析涉及的环节与任务众多，包含数据获取、数据存储、数据传输、数据操作、

数据分析、数据展示等，要求也各不相同。Tableau 并不是万能药，在有些数据处理工作上，还需要与其他工具和技术结合，才能够实现最大的成效。

埃森哲数据分析团队基于自身的项目经验与使用体会，结合实际的应用案例，总结了 Tableau 的功能特点与使用技巧，写作了本书，并且后续还会继续发展完善。对于希望或正在建设数据分析能力、开展数据分析业务的企业、组织或个人，本书可以提供一个不一样的选择与体验，拓宽对数据分析业务和工具的认识，帮助感兴趣的使用者更好地应用 Tableau 工具开展工作。

王靖

埃森哲大中华区董事总经理

前　言

从 2004 年最早接触电力市场分析开始，十余年来，我们先后参与了华东电力市场分析能力、华北电网市场分析体系、国家电网运营监测（控）中心等方面的建设工作，既深刻认识到业务监测与数据分析工作的价值，又充分感受到监测能力与分析体系建设的艰难与不易。我们看到了，固然有很多不错的值得借鉴的成功经验，但更注意到失败的实践比比皆是，由此造成了相当大的投资浪费，并且进一步限制了数据分析与已有业务的融合，影响了企业管理与经营绩效的提升。

自 2008 年以来，智能电网兴起，电网企业信息化建设日渐完善，我们正在步入大数据时代，数量规模急剧膨胀，数据种类日益繁多，更新速度不断加快。一方面，海量的数据中隐藏了诸多有价值的信息，通过有效的数据分析与挖掘，产生合理的业务见解，可以直接帮助企业提升竞争力；另一方面，这又增加了数据监测、分析工作的难度，传统的分析理念、分析工具更加难以适应新的发展形势。为此，亟需提出新的手段和方法，于是敏捷分析方法与工具应运而生。

传统商务智能（BI）实施周期动辄数月甚至长达一两年，严格地以清晰的业务需求为前提，且受限于传统 BI 工具，数据细节无法有效动态挖掘，它越来越不能满足企业的实际需要。数据监测与分析的路径和方向众多，不断尝试、往复迭代是发现问题、形成分析结论的必由之路，在实际做分析之前很难预先设计出来，分析思路与分析过程相辅相承；并且企业的实际业务发展速度也越来越快，所以根据所谓分析需求实施的 BI 往往开发出来就过气了，鲜有成功案例。

敏捷分析或商务智能不讲求大规模的数据建模，直接利用轻型分析应用，针对各类数据快速进行监测、分析业务探索，是一种想做相结合、过程结果持续循环的新型工作方式。敏捷分析很好地适应了监测、分析业务需求快速变化的特点，它的显著特征是轻量、快速、灵活，便于开展动态业务分析。当然，好的方法还要有好的软件工具作为支撑，Tableau 是支持敏捷分析的最为有效的工具之一，它在 Gartner 魔力四象限中的排名不断提升，目前在同类产品中排名第一，市场份额逐年翻倍。

有感于此，我们萌发了编写本书的念头，希望能够总结已有的经验，让更多的人掌握敏捷分析方法与工具，促进数据监测分析的发展。书中我们以目前电力行业已有的监测、分析业务实践为基础，全篇以丰富的电力行业实际案例贯穿始终，对各类方法、技术进行详细说明，包括数据连接与管理、初级与高级可视化分析、地图分析、高级数据操作、统计分析、分析图表整合与分析成果共享等主要内容，方便大家快速掌握敏捷分析方法与技术。同时，我们将 Tableau 的核心功能融入其中，详细介绍了该软件的数据连接与编辑、图形展示与编辑功能，阐述了如何与 R 等工具进行集成，如何在服务器上进行发布管理等内容，以方便读者快速学习 Tableau 的功能，

Tableau 产品自身也在快速更新升级，本书内容是基于最新发布的 9.1 版本编写的。数据源可以从图灵社区本书主页 http://www.ituring.com.cn/book/1736 免费下载使用。

本书是我们在日常工作之余，牺牲休息时间，历时半年之久增删补充而成，相信会给大家理解和应用敏捷分析方法和工具带来帮助。在即将付梓之时，我们特别感谢朱治中博士在写作资源上的全力支持，以及在写作上给予的方向性指导和建议；也特别感谢项目组同事石晶、杨宣华、张泽中、周跃、杨馨慧、楼琦瑶、陈寅迪等在写作过程中给予的无私帮助和支持；感谢 Tableau 大中华区同事 Alex、Julia、Raymond 等，他们给予的关心与鼓励是我们坚持前行的动力。同时我们要感谢图灵公司总经理武卫东对“数据分析创造价值”的认同，才让本书得以出版；感谢责任编辑王军花、策划编辑张霞不厌其烦、逐字逐图的校核，才让原稿得以成书；感谢封面设计师刘超的精彩创意，以及设计师黄智慧给出的关键建议，才让本书清爽面市。

虽然经过多次审核检查，书中难免还会存在一些错误与不足，恳请读者批评指正，并欢迎通过邮箱 fengxishi_15@163.com 与我们进行交流。

目　　录

第 1 章 Tableau入门

本章首先介绍了敏捷商务智能的发展进程和应用前景，然后介绍了作为数据可视化明星的Tableau软件的发展背景和独有特性，接着介绍了Tableau软件的丰富产品体系，最后带用户熟悉各种Tableau工作环境并了解Tableau的文件组织和管理方法。

如果不是初次接触Tableau，可以跳过本章，从第2章开始学习。

1.1 敏捷商务智能

当今社会，**商务智能**（Business Intelligence，BI）已被广泛地应用于各行各业，并在辅助企业的分析决策中扮演着举足轻重的角色。但随着企业数据量不断膨胀，IT环境日益复杂，业务需求灵活多变，以及信息实时性要求不断提高等，传统BI部署方案的弊端越来越明显。

传统BI系统架构的底层是数据源，中层是ETL（抽取、转换、装载），上层的数据仓库形成DWD（明细数据层）。业务人员进行数据分析，需IT人员根据预先定义好的分析需求，对明细数据进行汇总、建模形成DWA（汇总数据层），并通过前端展现工具制作报表，业务人员再在前端查看这些预生成的报表结果辅助分析。典型的传统BI系统架构图如图1-1所示。

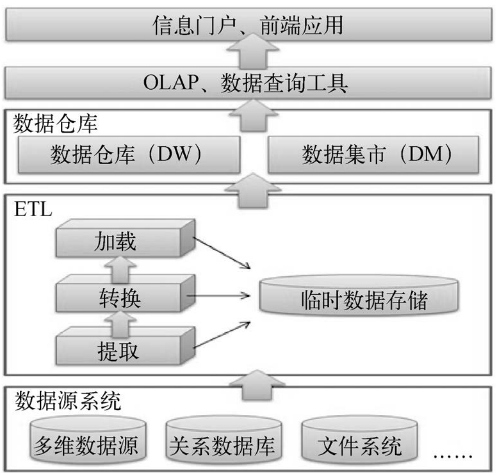


图1-1　典型的传统BI系统架构图

这种架构在处理企业数据时存在许多弊端，如下所示。

- 传统BI的开发难度较高，上线周期长，建好一套BI系统的开发周期长达几个月甚至半年；
- 传统BI的系统架构笨重，不能灵活地响应业务需求的变化，在面对需求变更时，调整周期长，无法提供自助式BI服务；
- 运维成本较高，因企业的分析需求不断变化，传统BI系统对变更需求的持续支持需投入大量的人力和财力；
- 整个BI系统一般由多个产品组成，总成本比较昂贵。

随着技术的更新和发展，以及企业对BI系统的轻便性、灵活性的要求日趋强烈，新一代敏捷BI应运而生。敏捷BI与传统BI相比，可以通过更低的成本、更短的上线周期、更快速地响应企业分析需求的变化，从而帮助企业及时洞察到数据的含义和价值。

敏捷BI具有以下优点。

- 直接把数据装载到内存数据集市中，无需预生成Cube，业务用户就可以通过自服务的方式直接在前端与数据进行交互分析，大大缩短了系统的上线周期；
- 基于细节数据，用户可以实现明细数据级的多维度探索式分析，而不再是仅能利用现有的分析模型，提高了对灵活多样的分析需求的支持度；
- 整个敏捷BI系统往往只需要一个产品即可实现，成本较传统BI系统低了很多。

1.2　数据可视化明星 Tableau

数据可视化是指借助于图形化的手段，清晰、有效地传达与沟通信息。随着信息技术的不断发展，当今社会已步入大数据时代，如何帮助企业在海量数据中快速获取重要信息应对市场变化，已成为企业亟需解决的难题。

Tableau是一款定位于数据可视化敏捷开发和实现的商务智能展现工具，可以用来实现交互的、可视化的分析和仪表板应用，从而帮助企业快速地认识和理解数据，以应对不断变化的市场环境与挑战。数据可视化让枯燥的数据以简单友好的图表形式展现出来，是一种最为直观有效的分析方式。无需过多的技术基础，任何个人、企业都可以轻松学会Tableau，并运用其可视化功能对数据进行处理和展示，从而更好地进行数据分析工作。

数据可视化技术是Tableau的核心，主要包括以下两个方面。

- **独创的VizQL数据库。**Tableau的初创合伙人是来自斯坦福大学的数据科学家，他们为了实现卓越的可视化数据获取与后期处理，并没有像普通数据分析类软件那样简单地调用和整合现行主流的关系型数据库，而是革命性地进行了大尺度的创新，独创了VizQL数据库。
- **用户体验良好且易用的表现形式。**Tableau提供了一个非常新颖而易用的使用界面，使得处理规模巨大的、多维的数据时，可以即时地从不同角度和设置看到数据所呈现出的规律。Tableau通过数据可视化技术，使得数据挖掘变得平民化，而其自动生成和展现出的图表，也丝毫不逊色于互联网美术编辑的水平。正是这个特点奠定了其广泛的用户基础（用户总数年均增长126%），带来了高续订率（90%的用户选择续订其服务）。

1.3　Tableau 的主要特性

Tableau作为轻量级可视化BI工具的优秀代表，在Gartner（高德纳）2015年2月公布的商业智能和分析平台魔力象限报告中，连续第三次蝉联领先者殊荣（如图1-2所示）。Gartner认为“Tableau在简单易用性方面排在现有所有BI工具的首位”。德国电子商务网站的数据科学家Lucie Salwiczek也认为：“不管是制作报表，还是深入挖掘数据并进行分析，只需要Tableau这样一个工具就够了。”



图1-2　Gartner商务智能及分析平台魔力象限图（2015年2月）

Tableau之所以在业界有如此出色的表现，在于以下几个方面的主要特性。

1. 极速高效

传统BI通过ETL过程处理数据，数据分析往往会延迟一段时间。而Tableau通过内存数据引擎，不但可以直接查询外部数据库，还可以动态地从数据仓库抽取数据，实时更新连接数据，大大提高了数据访问和查询的效率。

此外，用户通过拖放数据列就可以由VizQL转化成查询语句，从而快速改变分析内容；单击就可以突出变亮显示，并可随时下钻或上卷查看数据；添加一个筛选器、创建一个组或分层结构

就可变换一个分析角度，实现真正灵活、高效的即时分析。

2. 简单易用

简单易用是Tableau非常重要的一个特性。Tableau提供了非常友好的可视化界面，用户通过轻点鼠标和简单拖放，就可以迅速创建出智能、精美、直观和具有强交互性的报表和仪表盘。Tableau的简单易用性具体体现在以下两个方面。

- **易学，不需要技术背景和统计知识。**使用者不需要IT背景，也不需要统计知识，只通过拖放和点击（点选）的方式就可以创建出精美、交互式仪表盘。帮助迅速发现数据中的异常点，对异常点进行明细钻取，还可以实现异常点的深入分析，定位异常原因。
- **操作极其简单。**对于传统BI工具，业务人员和管理人员主要依赖IT人员定制数据报表和仪表盘，并且需要花费大量时间与IT人员沟通需求、设计报表样式，而只有少量时间真正用于数据分析。Tableau具有友好且直观的拖放界面，操作上类似Excel数据透视表，即学即会即用，IT人员只需将数据准备好，并开放数据权限，业务人员或管理人员就可以连接数据源自己来做分析。

3. 可连接多种数据源，轻松实现数据融合

在很多情况下，用户想要展示的信息分散在多个数据源中，有的存在于文件中，有的可能存放在数据库服务器上。Tableau允许从多个数据源访问数据，包括带分隔符的文本文件、Excel文件、SQL数据库、Oracle数据库和多维数据库等。Tableau也允许用户查看多个数据源，在不同的数据源间来回切换分析，并允许用户把多个不同数据源结合起来使用。

此外，Tableau还允许在使用关系数据库或文本文件时，通过创建联接（支持多种不同联接类型，如左侧联接、右侧联接和内部联接等）来组合多个表或文件中存在的数据，以允许分析相互有关系的数据。

4. 高效接口集成，具有良好可扩展性，提升数据分析能力

Tableau提供多种应用编程接口，包括数据提取接口、页面集成接口和高级数据分析接口，具体包括以下几个。

- **数据提取API。**Tableau可以连接使用多种格式数据源，但由于业务的复杂性，数据源的格式多种多样，Tableau所支持的数据源格式不可能面面俱到。为此，Tableau提供了数据提取API，使用它们可以在C、C++、Java或Python中创建用于访问和处理数据的程序，然后使用这样的程序创建Tableau 数据提取（.tde）文件。
- **JavaScript API。**通过JavaScript API，可以把通过Tableau制作的报表和仪表盘嵌入到已有的企业信息化系统或企业商务智能平台中，实现与页面和交互的集成。
- **与数据分析工具R的集成接口。**R 是一种用于统计分析和预测建模分析的开源软件编程语言和软件环境，具有非常强大的数据处理、统计分析和预测建模能力。Tableau8.1之后的版本，支持与R的脚本集成，大大提升了Tableau在数据处理和高级分析方面的能力。

1.4 Tableau 的产品体系

Tableau的产品体系非常丰富，不仅包括制作报表、视图和仪表板的桌面端设计和分析工具，还包括适用于企业部署的Tableau服务器产品，还有适用于网页上创建和分享数据可视化内容的完全免费服务产品Tableau Public。

1. Tableau Desktop

Tableau Desktop是设计和创建美观的视图与仪表板、实现快捷数据分析功能的桌面端分析工具，包括 Tableau Desktop Personal（个人版）和 Tableau Desktop Professional（专业版）两个版本，支持Windows和Mac操作系统。

Tableau个人版仅允许连接到文件和本地数据源，分析成果可以发布为图片、PDF和Tableau Reader等格式；而Tableau专业版除了具备个人版的全部功能外，支持的数据源更加丰富，能够连接到几乎所有格式的数据和数据库系统，包括以ODBC方式新建数据源库，分析成果还可以发布到企业或个人的Tableau 服务器、Tableau Online服务器和Tableau Public服务器上，实现移动办公。因此，专业版比个人版更加通用，但个人版的价格相对专业版也便宜不少。

2. Tableau Server

Tableau Server是一款商业智能应用程序，用于发布和管理Tableau Desktop 制作的报表，也可以发布和管理数据源，如自动刷新发布到Server上的数据提取。Tableau Server是基于浏览器的分析技术，非常适用于企业范围内的部署，当工作簿做好并发布到Tableau Server上后，用户可以通过浏览器或移动终端设备，查看工作簿的内容并与之交互。

Tableau Server可控制对数据连接的访问权限，并允许针对工作簿、仪表板甚至用户设置来设置不同安全级别的访问权限。通过Tableau Server提供的访问接口，用户可以搜索和组织工作簿，还可以在仪表板上添加批注，与同事分享数据见解，实现在线互动。利用Tableau Server提供的订阅功能，当允许访问的工作簿版本有更新时，用户可以接收到邮件通知。

3. Tableau Online

Tableau Online 针对云分析而建立，是 Tableau Server 的一种托管版本，省去硬件部署、维护及软件安装的时间与成本，提供的功能与Tableau Server没有区别，按每人每年的方式付费使用。

4. Tableau Mobile

Tableau Mobile是基于iOS和Android平台移动端应用程序。用户可通过iPad、Android设备或移动浏览器，来查看发布到Tableau Server或Tableau Online上的工作簿，并可进行简单的编辑和导出操作。

5. Tableau Public

Tableau Public 是一款免费的桌面应用程序，用户可以连接Tableau Public服务器上的数据，设计和创建自己的工作表、仪表板和工作簿，并把成果保存到大众皆可访问的Tableau Public服务器上（不可以把成果保存到本地电脑上）。Tableau Public使用的数据和创建的工作簿都是公开的，任何人都可以与其互动并可随意下载，还可以根据你的数据创建自己的工作簿。

6. Tableau Reader

Tableau Reader是一个免费的桌面应用程序，可以用来打开和查看打包工作簿文件（.twbx），也可以与工作簿中的视图和仪表板进行交互操作，如筛选、排序、向下钻取和查看数据明细等。打包工作簿文件可以通过Tableau Desktop创建和发布，也可以从Tableau Public服务器下载。用户无法使用Tableau Reader创建工作表和仪表板，也无法改变工作簿的设计和布局。

说明 利用Tableau Public连接数据时，对数据源、数据文件大小和长度都有一定限制：仅包括Excel、Access和多种文本文件格式，对单个数据文件的行数限制为10万行，对数据的存储空间限定在50MB以内。Tableau Public Premium是Tableau Public的高级产品，主要提供给某些组织使用，它提供了更大的数据处理能力和允许隐藏底层数据的功能。

1.5 Tableau 的工作区

在首次进入Tableau或打开Tableau但没有指定工作簿时，会显示“开始页面”，其中包含了最近使用的工作簿、已保存的数据连接、示例工作簿和其他一些入门资源，这些内容将帮助初学者快速入门，由于比较简单直观，本节对开始页面不作介绍。要开始构建视图并分析数据，还需要先进入“新建数据源”页面，将 Tableau 连接到一个或多个数据源，有关连接数据页面的详细介绍可以参见第3章。

Tableau工作区是制作视图、设计仪表板、生成故事、发布和共享工作簿的工作环境，包括工作表工作区、仪表板工作区和故事工作区，也包括公共菜单栏和工具栏。在正式介绍各工作区环境之前，首先需要了解以下几个基本概念。

- **工作表**（work sheet）：又称为视图（visualization），是可视化分析的最基本单元。
- **仪表板**（dashboard）：是多个工作表和一些对象（如图像、文本、网页和空白等）的组合，可以按照一定方式对其进行组织和布局，以便揭示数据关系和内涵。
- **故事**（story）：是按顺序排列的工作表或仪表板的集合，故事中各个单独的工作表或仪表板称为“故事点”。可以使用创建的故事，向用户叙述某些事实，或者以故事方式揭示各种事实之间的上下文或事件发展的关系。
- **工作簿**（workbook）：包含一个或多个工作表，以及一个或多个仪表板和故事，是用户在Tableau中工作成果的容器。用户可以把工作成果组织、保存或发布为工作簿，以便共享和存储。

1.5.1 工作表工作区

工作表工作区（见图1-3）包含菜单、工具栏、数据窗口、含有功能区和图例的卡，可以在工作表工作区中通过将字段拖放到功能区上来生成数据视图（工作表工作区仅用于创建单个视图）。在Tableau中连接数据之后，即可进入工作表工作区。

工作表工作区中的主要部件如下。

- **数据窗口**。数据窗口位于工作表工作区的左侧。可以通过单击数据窗口右上角的最小化按钮来隐藏和显示数据窗口，这样数据窗口会折叠到工作区底部，再次单击最小化按钮可显示数据窗口。通过单击，然后在文本框中键入内容，可在数据窗口中搜索字段。通过单击，可以查看数据。数据窗口由数据源窗口、维度窗口、度量窗口、集窗口和参数窗口等组成。

  数据源窗口：包括当前使用的数据源及其他可用的数据源。请参见第3章各节内容。

  维度窗口：包含诸如文本和日期等类别数据的字段。请参见2.2.1节数据角色中的维度。

  度量窗口：包含可以聚合的数字的字段。请参见2.1.1节数据角色中的度量。

  集窗口：定义的对象数据的子集，只有创建了集，此窗口才可见。请参见6.3节。

  参数窗口：可替换计算字段和筛选器中的常量值的动态占位符，只有创建了参数，此窗口才可见。请参见6.4节参数。

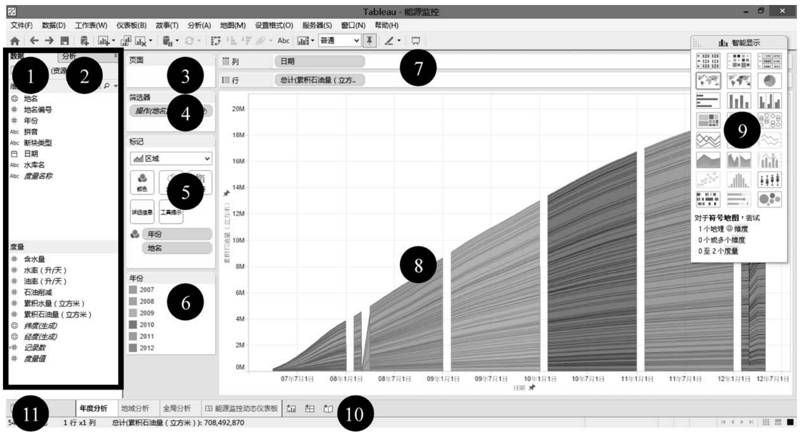


图1-3 工作表工作区

- **分析窗口**。将菜单中常用的分析功能进行了整合，方便快速使用，主要包括汇总、模型和自定义3个窗口。

  汇总窗口：提供常用的参考线、参考区间及其他分析功能，包括常量线、平均线、含四分位点的中值、盒须图和合计等，可直接拖放到视图中应用。

  模型窗口：提供常用的分析模型，包括含95% CI的平均值、趋势线和预测。

  自定义窗口：提供参考线 、参考区间、分布区间和盒须图的快捷使用，请参见6.7节。

- **页面卡**。可在此功能区上基于某个维度的成员或某个度量的值将一个视图拆分为多个视图，请参见2.3.4节页面。
- **筛选器卡**。指定要包含和排除的数据，所有经过筛选的字段都显示在筛选器卡上，请参见2.3.3节筛选器。
- **标记卡**。控制视图中的标记属性，包括一个标记类型选择器，可以在其中指定标记类型

（例如，条、线、区域等）。此外，还包含颜色、大小、标签、文本、详细信息、工具提示、形状、路径和角度等控件，这些控件的可用性取决于视图中的字段和标记类型。请参见2.3.2标记卡。

- **颜色图例。**包含视图中颜色的图例，仅当颜色上至少有一个字段时才可用。同理，也可以添加形状图例、尺寸图例和地图图例。
- **行功能区和列功能区。**行功能区用于创建行，列功能区用于创建列，可以将任意数量的字段放置在这两个功能区上。
- **工作表视图区。**创建和显示视图的区域，一个视图就是行和列的集合，由以下组件组成：标题、轴、区、单元格和标记。除这些内容外，还可以选择显示标题、说明、字段标签、摘要和图例等。
- **智能显示。**通过智能显示，可以基于视图中已经使用的字段以及在数据窗口中选择的任何字段来创建视图。Tableau会自动评估选定的字段，然后在智能显示中突出显示与数据最相符的可视化图表类型。
- **标签栏。**显示已经被创建的工作表、仪表板和故事的标签，或者通过标签栏上的新建工作表图标创建新工作表，或者通过标签栏上的新建仪表板图标创建新仪表板。
- **状态栏。**位于Tableau工作簿的底部。它显示菜单项说明以及有关当前视图的信息。可以通过选择“窗口”➤“显示状态栏”来隐藏状态栏。有时Tableau 会在状态栏的右下角显示警告图标，以指示错误或警告。

1.5.2　仪表板工作区

仪表板工作区使用布局容器把工作表和一些像图片、文本、网页类型的对象按一定的布局方式组织在一起。在工作区页面单击新建仪表板图标，或者选择“仪表板”➤“新建仪表板”，打开仪表板工作区，仪表板窗口将替换工作表左侧的数据窗口。图1-4显示了Tableau中的仪表板工作区。

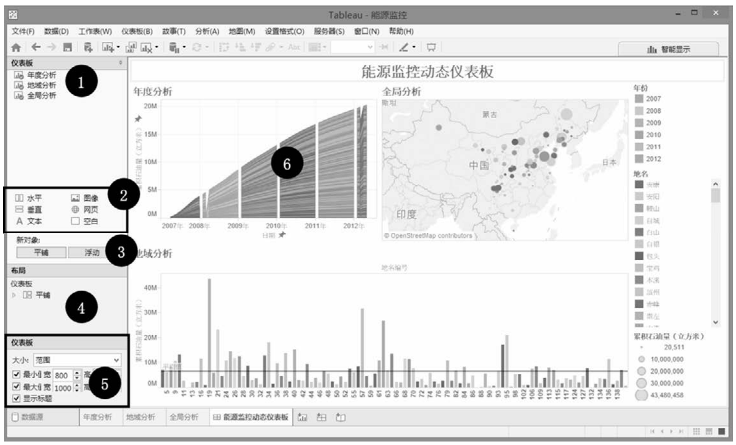


图1-4　仪表板工作区

仪表板工作区中的主要部件如下。

- **仪表板窗口**。列出了在当前工作簿中创建的所有工作表，可以选中工作表并将其从仪表板窗口拖至右侧的仪表板区域中，一个灰色阴影区域将指示出可以放置该工作表的各个位置。在将工作表添加至仪表板后，仪表板窗口中会用复选标记来标记该工作表。
- **仪表板对象窗口**。包含仪表板支持的对象，如文本、图像、网页和空白区域。从仪表板窗口拖放所需对象至右侧的仪表板窗口中，可以添加仪表板对象。
- **平铺和浮动**。决定了工作表和对象被拖放到仪表板后的效果和布局方式。默认情况下，仪表板使用平铺布局，这意味着每个工作表和对象都排列到一个分层网格中。可以将布局更改为浮动以允许视图和对象重叠。
- **布局窗口**。以树形结构显示当前仪表板中用到的所有工作表及对象的布局方式。
- **仪表板设置窗口**。设置创建的仪表板的大小，也可以设置是否显示仪表板标题。仪表板的大小可以从预定义的大小中选择一个，或以像素为单位设置自定义大小。
- **仪表板视图区**。是创建和调整仪表板的工作区域，可以添加工作表及各类对象。

1.5.3 故事工作区

故事是Tableau 8.2之后新增的特性，一般将故事用作演示工具，按顺序排列视图或仪表板。选择“故事”➤“新建故事”，或者单击工具栏上的新建工作表按钮，然后选择“新建故事”。故事工作区与创建工作表和仪表板的工作区有很大区别，如图1-5所示。



图1-5 故事工作区

故事工作区中的主要部件如下。

- **仪表板和工作表窗口**。显示在当前工作簿中创建的视图和仪表板的列表，将其中的一个视图或仪表板拖到故事区域（导航框下方），即可创建故事点，单击可快速跳转至所在

的视图或仪表板。
- **说明**。说明是可以添加到故事点中的一种特殊类型的注释。若要添加说明，只需双击此处。可以向一个故事点添加任何数量的说明，放置在故事中的任意所需位置上。
- **导航器设置**。设置是否显示导航框中的后退/前进按钮。
- **故事设置窗口**。设置创建的故事的大小，也可以设置是否显示故事标题。故事的大小可以从预定义的大小中选择一个，或以像素为单位设置自定义大小。
- **导航框**。用户进行故事点导航的窗口，可以利用左侧或右侧的按钮顺序切换故事点，也可以直接单击故事点进行切换。
- **新空白点按钮**。单击此按钮可以创建新故事点，使其与原来的故事点有所不同。
- **复制按钮**。可以将当前故事点用作新故事点的起点。
- **说明框**。通过说明为故事点或者故事点中的视图或仪表板添加的注释文本框。
- **故事视图区**。是创建故事的工作区域，可以添加工作表、仪表板或者说明框对象。

1.5.4 菜单栏和工具栏

除了工作表、仪表板和故事工作区，Tableau工作区环境还包括公共的菜单栏和工具栏。无论在哪个工作区环境下，菜单栏和工具栏都存在于工作区的顶部。

1. 菜单栏

菜单栏包括文件、数据、工作表和仪表板等菜单，每个菜单下都包含很多菜单选项，下面将主要介绍各菜单中的常用功能选项。
- **文件菜单**。像任何文件菜单一样，该菜单包括打开、保存和另存为等功能。文件菜单中最常用的功能是“打印为PDF…”选项，它允许把工作表或仪表板导出为PDF。“导出打包工作簿…”选项也非常常用，它允许把当前的工作簿以打包形式导出，想了解这种导出方式与其他导出方式有何不同，请参见1.6节。如果记不清文件存储位置，或者想要改变文件的缺省存储位置，可以使用文件菜单中的“存储库位置…”选项来查看文件存储位置和改变文件的缺省存储位置。
- **数据菜单**。数据菜单中的“粘贴数据…”选项非常方便，如果在网页上发现了一些Tableau的数据，并且想要使用Tableau进行分析，可以从网页上复制下来，然后使用此选项把数据导入到Tableau中进行分析。一旦数据被粘贴，Tableau将从Windows粘贴板中复制这些数据，并在数据窗口中增加一个数据源。“编辑关系…”选项在数据融合时使用，它可以用于创建或修改当前数据源关联关系，并且如果两个不同数据源中的字段名不相同，此选项非常有用，它允许明确地定义相关的字段。
- **工作表菜单**。工作表菜单中有几个常用的功能，如“导出”选项和“复制”选项。其中“导出”选项允许把工作表导出为一个图像、一个Excel交叉表或者Access数据库文件（.mdb）。而使用“复制”选项中的“复制为交叉表”选项会创建一个当前工作表的交叉表版本，并把它存放在一个新的工作表中。

- **仪表板菜单**。此菜单中的选项只有在仪表板工作区环境下可用，第9章详细介绍了每个选项的使用方法。
- **故事菜单**。此菜单中的选项只有在故事工作区环境下可用，可以利用其中的选项新建故事，利用“设置格式”选项设置故事的背景、标题和说明，还可以利用“导出图像...”选项把当前故事导出为图像。
- **分析菜单**。在熟悉了Tableau的基本视图创建方法后，可以使用分析菜单中的一些选项来创建高级视图，或者利用它们来调整Tableau中的一些缺省行为，如利用其中的“聚合度量”选项来控制对字段的聚合或解聚，也可以利用“创建计算字段...”和“编辑计算字段”选项创建当前数据源中不存在的字段。第6章“高级数据操作”和第7章“高级可视化分析”中将频繁使用分析菜单中的各个选项。分析菜单在故事工作区环境下不可见，在仪表板工作区环境下仅部分功能可用。
- **地图菜单**。地图菜单中的“地图选项...”里的“样式”可以更改地图颜色配色方案，如选择普通、灰色或者黑色地图样式，也可以使用“地图选项...”中的“冲蚀”滑块控制背景地图的强度或亮度，滑块向右移得越远，地图背景就越模糊。地图菜单中的“地理编码”选项可以导入自定义地理编码文件，绘制自定义地图。第5章“地图分析”详细介绍了地图菜单中各个选项的使用方法。
- **设置格式菜单**。设置格式菜单很少使用，因为在视图或仪表板上的某些特定区域单击右键可以更快捷地调整格式。但有些设置格式菜单中的选项通过快捷键方式无法实现，例如想要修改一个交叉表中单元格的尺寸，只能利用设置格式菜单中的“单元格大小”选项来调整；如果不喜欢当前工作簿的默认主题风格，只能利用“工作簿主题”选项来切换至其他两个子选项“现代”或“古典”。
- **服务器菜单**。如果想要把工作成果发布到大众皆可访问的公共服务器Tableau Public上，或者从上面下载或打开工作簿，可以使用服务器菜单中的“Tableau Public”选项。如果需要登录到Tableau服务器，或者需要把工作成果发布到Tableau服务器上，需要使用服务器菜单中的“登录”选项。详细的服务器连接和配置方法，请参见第11章“Tableau Server简介”。
- **窗口菜单**。如果工作簿很大，其中包含了很多工作表，并且想要把其中某个工作表共享给别人，可以使用窗口菜单中的“书签”选项创建一个书签文件（.tbm），还可以通过窗口菜单中的其他选项，来决定显示或隐藏工具栏、状态栏和边条。
- **帮助菜单**。最右侧的帮助菜单可以让用户直接连接到Tableau的在线帮助文档、培训视频、示例工作簿和示例库，也可以设置工作区语言。此外，如果加载仪表板时比较缓慢，可以使用“设置和性能”选项中的子选项“启动性能记录”激活Tableau的性能分析工具，优化加载过程。

2. 工具栏

工具栏包含“新建数据源”“新建工作表”和“保存”等命令。另外，该工具栏还包含“排序”“分组”和“突出显示”等分析和导航工具。通过选择“窗口”➤“显示工具栏”可隐藏或

显示工具栏。工具栏有助于快速访问常用工具和操作，其中有些命令仅对工作表工作区有效，有些命令仅对仪表板工作区有效，有些命令仅对故事工作区有效。表1-1详细解释了每个工具栏按钮的主要功能。

表1-1　工具栏说明表

图　标	说　明
	撤销：反转工作簿中的最新操作。可以无限次撤销，返回到上次打开的工作簿，即使是在保存之后也可撤销
	重做：重复使用“撤销”按钮反转的最后一个操作，可以重做无限次
	保存：保存对工作簿进行的更改
	新建数据源：打开“新建数据源”页，可以在其中创建新连接，或者从存储库中打开已保存的连接
	新建工作表：新建空白工作表。使用下拉菜单可创建新工作表、仪表板或故事
	复制工作表：创建含有与当前工作表完全相同的视图的新工作表
	清除：清除当前工作表。使用下拉菜单清除视图的特定部分，如筛选器、格式设置、大小调整和轴范围
	自动更新：控制进行更改后Tableau是否自动更新视图。使用下拉列表来自动更新整个工作表或只使用快速筛选器
	运行更新：运行手动数据查询，以便在关闭自动更新后用所做的更改对视图进行更新。使用下拉菜单来更新整个工作表或只使用快速筛选器
	交换：交换行功能区和列功能区上的字段。每次按此按钮，都会交换“隐藏空行”和“隐藏空列”设置
	升序排序：根据视图中的度量，以所选字段的升序来应用排序
	降序排序：根据视图中的度量，以所选字段的降序来应用排序
	成员分组：通过组合所选值来创建组。选择多个维度时，使用下拉菜单指定是对特定维度进行分组，还是对所有维度进行分组
Abc	显示标记标签：在显示和隐藏当前工作表的标记标签之间切换
	演示模式：在显示和隐藏视图（即功能区、工具栏、数据窗口）之外的所有内容之间切换
	查看卡：显示和隐藏工作表中的特定卡。在下拉菜单上选择要隐藏或显示的每个卡
Normal	适合选择器：指定在应用程序窗口中调整视图大小的方式。可选择“标准适合”“适合宽度”“适合高度”或“整个视图”
	固定轴：在仅显示特定范围的锁定轴以及基于视图中的最小值和最大值调整范围的动态轴之间切换
	突出显示：启用所选工作表的突出显示。使用下拉菜单中的选项定义突出显示值的方式

1.6　Tableau 的文件管理

可以使用多种不同的Tableau文件类型，如工作簿、打包工作簿、数据提取、数据源和书签等，来保存和共享工作成果和数据源（见表1-2）。

表1-2　Tableau文件类型表

文件类型	大　小	使用场景	内　容
Tableau工作簿（.twb）	小	Tableau缺省保存工作的方式	可视化内容，但无源数据
Tableau打包工作簿（.twbx）	可能非常大	与无法访问数据源的用户分享工作	创建工作簿的所有信息和资源

(续)

文件类型	大　小	使用场景	内　容
Tableau数据源（.tds）	极小	频繁使用的数据源	包含新建数据源所需的信息，如数据源类型和数据源连接信息，数据源上的字段属性以及在数据源上创建的组、集和计算字段等
Tableau数据源（.tdsx）	小	频繁使用的数据源	包含数据源（.tds）文件中的所有信息以及任何本地文件数据源（Excel、Access、文本和数据提取）
Tableau书签（.tbm）	通常很小	工作簿间分享工作表时使用	如果原始工作簿是一个打包工作簿，创建的书签就包含可视化内容和书签
Tableau数据提取（.tde）	可能非常大	提高数据库性能	部分或整个数据源的一个本地副本

下面对常用的文件类型分别进行介绍。

- **Tableau工作簿（.twb）**：将所有工作表及其连接信息保存在工作簿文件中，不包括数据。
- **打包工作簿（.twbx）**：打包工作簿是一个 zip 文件，保存所有工作表、连接信息以及任何本地资源（如本地文件数据源、背景图片、自定义地理编码等）。这种格式最适合对工作进行打包以便与不能访问该数据的其他人共享。
- **Tableau数据源（.tds）**：Tableau 数据源文件具有 .tds 文件扩展名。数据源文件是快速连接经常使用的数据源的快捷方式。数据源文件不包含实际数据，只包含新建数据源所必需的信息以及在数据窗口中所做的修改，例如默认属性、计算字段、组、集等。
- **Tableau数据源（.tdsx）**：如果连接的数据源不是本地数据源，tdsx文件与tds文件没有区别。如果连接的数据源是本地数据源，数据源（.tdsx）不但包含数据源（.tds）文件中的所有信息，还包括本地文件数据源（Excel、Access、文本和数据提取）。
- **Tableau书签（.tbm）**：书签包含单个工作表，是快速分享所做工作的简便方式。
- **Tableau数据提取（.tde）**：Tableau 数据提取文件具有 .tde 文件扩展名。提取文件是部分或整个数据源的一个本地副本，可用于共享数据、脱机工作和提高数据库性能。

这些文件可保存在“我的 Tableau 存储库”目录中的关联文件夹中，该目录是在安装 Tableau 时在“我的文档”文件夹中自动创建的。工作文件也可保存在其他位置，如桌面上或网络目录中。

第 2 章

典型应用场景

2

简便、快速地创建视图和仪表板是Tableau最大的优点之一。本章将首先介绍Tableau的数据基础，然后通过一个案例展示Tableau创建、设计、保存视图和仪表板的基本方法和主要操作步骤。希望通过本章的学习，读者能够了解Tableau支持的数据角色和字段类型的概念，熟悉Tableau工作区中的各功能区的使用方法和操作技巧，最重要的目的在于利用Tableau快速创建基本的视图。

2.1　数据准备

本章案例的样本数据参见附录D.1，指标为售电量，统计周期为2014年01月～2014年06月，数据存储为Excel文件，结构见图2-1。共有6列变量，用电类别是对售电量市场的进一步细分，包括大工业、居民、非居民、商业等9类；当期值为统计周期对应时间的售电量；同期值为上一年相同月份的售电量；月度计划值为当月的计划值。打开Tableau桌面版，“新建数据源”➤“Microsoft Excel”，将该数据表导入到Tableau中，进入Tableau工作区，见图2-2。数据源中数据与Tableau中数据对应关系见图2-3。连接数据的详细方法参见第3章。

	A	B	C	D	E	F
1	省市	统计周期	用电类别	当期值	同期值	月度计划值
2	江苏	5/31/2014	大工业	2473806.39	1801205.88	1801205.88
3	江苏	6/30/2014	大工业	2459465.16	1815454.48	1815454.48
4	江苏	3/31/2014	大工业	2299171.73	1646656.54	1646656.54
5	江苏	4/30/2014	大工业	2266469.52	1659679.5	1659679.5
6	江苏	1/31/2014	大工业	2092388.83	1643401	1643401
7	江苏	2/28/2014	大工业	1897061.34	1062801.77	1062801.77
8	山东	5/31/2014	大工业	1607161.75	1303711	1303711
9	山东	4/30/2014	大工业	1594860.1	1313730	1313730
10	山东	6/30/2014	大工业	1565942.58	1302881	1302881
11	浙江	4/30/2014	大工业	1565738.35	1484657.43	1484657.43
12	浙江	6/30/2014	大工业	1564680.49	1488011.76	1488011.76
13	浙江	5/31/2014	大工业	1514825.81	1478757.19	1478757.19
14	山东	3/31/2014	大工业	1486366.42	1271142	1271142
15	浙江	1/31/2014	大工业	1387124.19	1412112.2	1412112.2
16	浙江	3/31/2014	大工业	1354949.99	1190055.11	1190055.11
17	山东	1/31/2014	大工业	1234932.57	1396797.5	1396797.5
18	山东	2/28/2014	大工业	1161511.46	1178342.07	1178342.07
19	河南	1/31/2014	趸售	1015447.12	976051	976051
20	河南	4/30/2014	大工业	975631.36	918596.54	918596.54
21	河南	5/31/2014	大工业	911216.46	897400.36	897400.36
22	河南	3/31/2014	大工业	907300.51	869560.82	869560.82
23	辽宁	1/31/2014	大工业	835727	856460	856460
24	辽宁	1/31/2014	居民	196408	207754	207754
25	辽宁	1/31/2014	非普工业	159107	169438	169438
26	河南	2/28/2014	趸售	869885.6	828267	828267
27	河南	6/30/2014	大工业	867164.57	920423.61	920423.61
28	山西	1/31/2014	大工业	849845.56	841130	841130

2014年各省售电量明细

图2-1　源数据

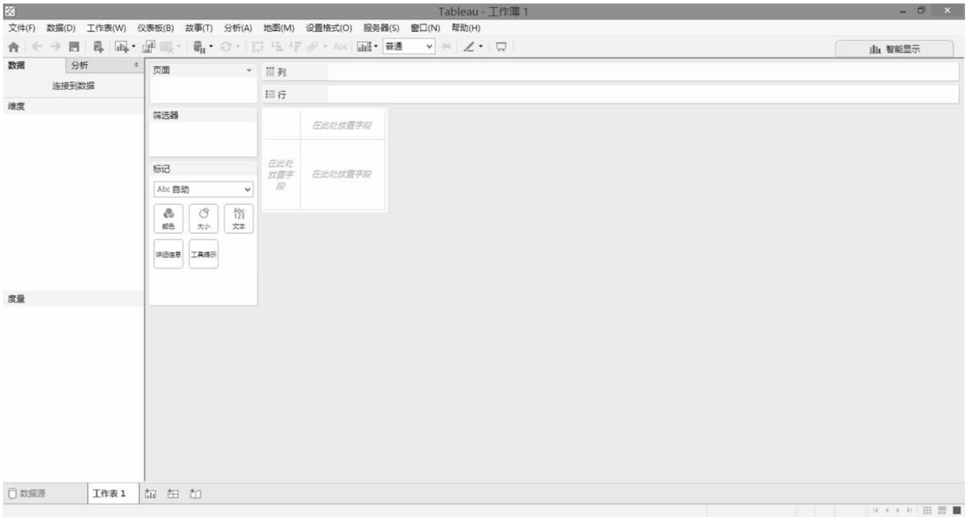


图2-2 工作区示意图

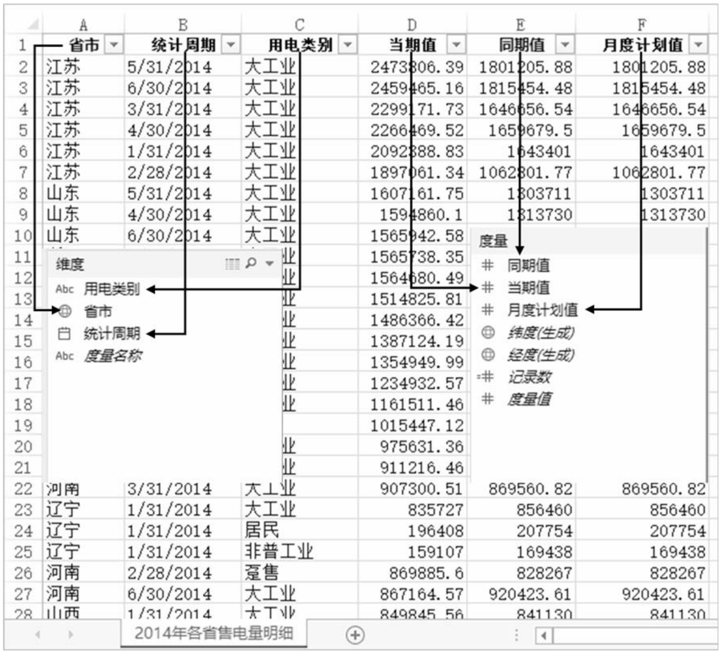


图2-3 数据对应关系

说明 本书案例数据均非真实数据，仅用以讲解Tableau的基本操作。

2.2　认识 Tableau 数据

本节主要介绍数据的角色（包含度量和维度，离散和连续）、字段类型和字段类型转换。

2.2.1　数据角色

Tableau连接数据后会将数据显示在工作区的左侧，我们称之为数据窗口，见图2-4框出部分。数据窗口的顶部是数据源窗口，其中显示的是连接到Tableau的数据源，Tableau支持连接多个数据源（详见第3章），数据源窗口的下方分别为维度窗口和度量窗口，分别用来显示导入的维度字段和度量字段（Tableau将数据表中的一列变量称为字段）。

维度和度量是Tableau的一种数据角色划分，离散和连续是另一种划分方式。Tableau功能区对不同数据角色操作处理方式是不同的，因此了解Tableau数据角色十分必要。

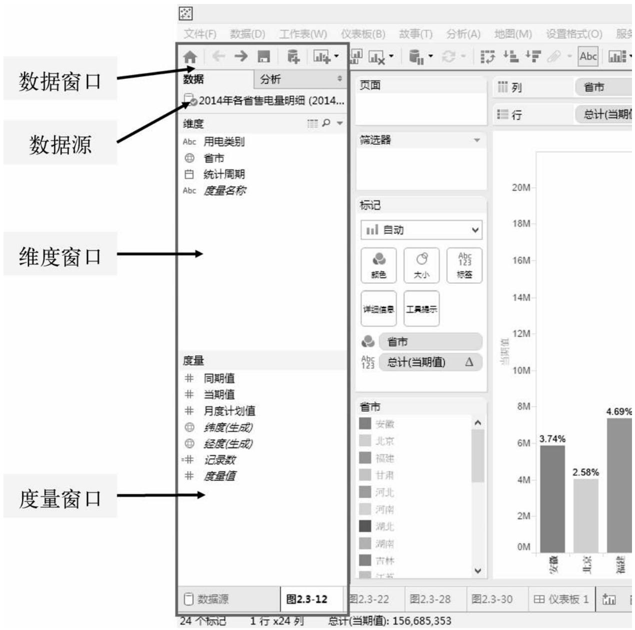


图2-4　数据窗口

1. 维度和度量

度量窗口显示的数据角色为度量，往往是数值字段，将其拖放到功能区时，Tableau默认会进行聚合运算，同时，视图区将产生相应的轴。维度窗口显示的数据角色为维度，往往是一些分类、时间方面的定性字段，将其拖放到功能区时，Tableau不会对其进行计算，而是对视图区进行分区，维度的内容显示为各区的标题。比如想展示各省售电量当期值，这时“省市”字段就是维度，“当期值”为度量，“当期值”将依据各省市分别进行“总计”聚合运算。

Tableau连接数据时会对各个字段进行评估，根据评估自动地将字段放入维度窗口或度量窗

口。通常Tableau的这种分配是正确的，但是有时也会出错。比如数据源中有员工工号字段时，工号由一串数字构成，连接数据源后，Tableau会将其自动分配到度量中。这种情况下，我们可以把工号从度量窗口拖放至维度窗口中，以调整数据的角色。例如将字段“当期值”转换为维度，只需将其拖放到维度窗口中即可。字段“当期值”前面的图标也会由绿色变为蓝色，见图2-5。

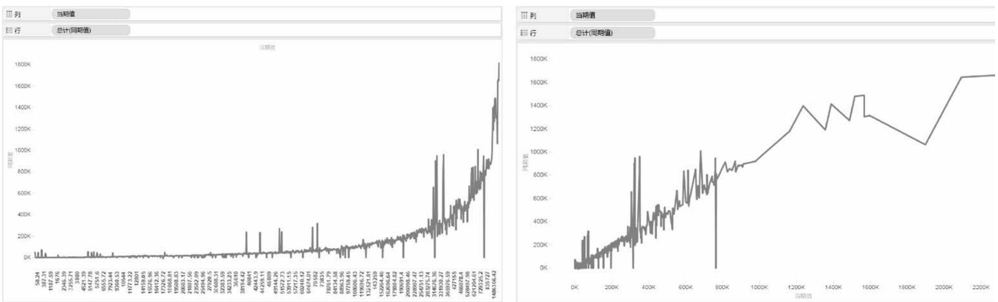


图2-5 维度与度量转换（另见彩插图2-5）

维度和度量字段有个明显的区别就是图标（类似 Abc 、 # ）颜色，维度是蓝色，度量是绿色。实际上在Tableau作图时这种颜色的区别贯穿始终，当我们创建视图拖放字段到行功能区或列功能区时，依然会保持相应的两种颜色。

2. 离散和连续

离散和连续是另一种数据角色分类，在Tableau中，蓝色是离散字段，绿色是连续字段。离散字段在行列功能区时总是在视图中显示为标题，而连续字段则在视图中显示为轴，如图2-6所示。



(a) 当期值为离散类型 (b) 当期值为连续类型

图2-6 离散和连续类型（另见彩插图2-6）

当期值为离散类型时，当期值中的每一个数字都是标题，字段颜色为蓝色。

当期值为连续类型时，下方出现的是一条轴，轴上是连续刻度，当期值是轴的标题，字段颜色为绿色。

离散和连续类型也可以相互转换，右键字段，在弹出框中就有“离散”和“连续”的选项，单击即可实现转换。

2.2.2　字段类型

数据窗口中各字段前如 Abc 、# 等符号是标示字段类型的图标。Tableau支持的数据类型见表2-1。

表2-1　Tableau支持的数据类型

显示的窗口	字段图标	字段类型	示　例	说　明
维度	Abc	文本	A,B,华北	
		日期	1/31/2014	日期的图标像日历，日期和时间的图标是日历加一个小时钟
		日期和时间	1/31/2014 08:31:42AM	
		地理值	北京，四川	用于地图
	T|F	布尔值	True/False	只有这两类值，仅限关系型数据源
度量	#	数字	1,12.1,30%	
	维度（生成）	地理编码		当数据中有地理类型名称时自动出现在度量中
	经度（生成）			

说明　=# 即数字标志符号前加个等号，表示这个字段不是原数据中的字段，而是Tableau自定义的一个数字型字段。同理，=Abc 是指Tableau自定义的一个字符串型字段。

2.2.3　字段类型转换

Tableau会自动对导入的数据分配字段类型，但有时自动分配的字段类型不是我们所希望的。由于字段类型对于视图的创建非常重要，因此一定要在创建视图前调整一些分配不规范的字段类型。

例如本案例图2-5中，我们发现字段“省市”和“统计周期”显示的字段类型都为字符串 Abc，而不是我们想要的地理和日期类型，这时就需要我们手动调整。调整方法为单击 Abc 省市 右侧小三角形（或者右键），在弹出的对话框中选择“地理角色”➤“省/市/自治区”，这时“省市”便成了地理字段，并且在选择后度量窗口会自动显示相应的经纬度字段，见图2-7。

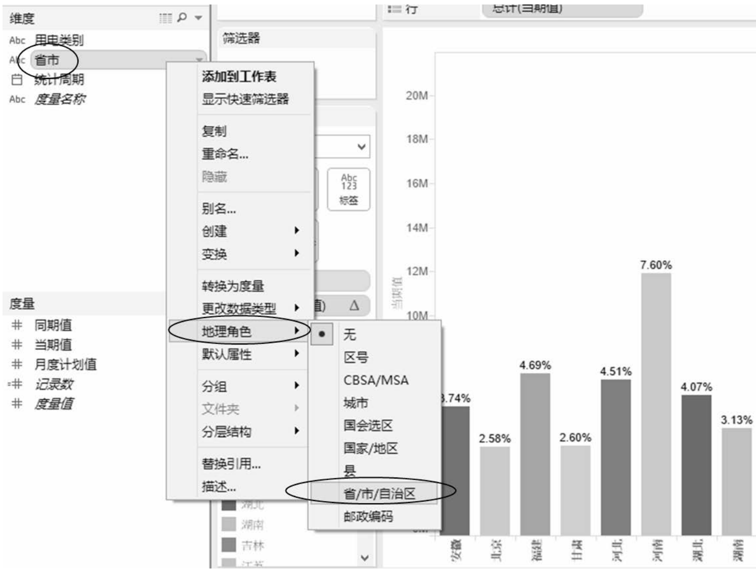


图2-7 更改字段类型

对于“统计周期”，同样选择“更改数据类型”➤“日期”即可。

可以发现在数据窗口有3个多余的字段：记录数、度量名称和度量值。实际上，每次新建数据源都会出现这3个字段，其中记录数是Tableau自动给每行观测值赋值为1，可用以计数。度量名称和度量值的使用详见2.3.6节。

2.3 创建视图

在对Tableau的数据有了基本的认识后，我们便可以创建Tableau视图了。一个完整的Tableau可视化产品由多个仪表板构成，每个仪表板由一个或多个视图（工作表）按照一定的布局方式构成，因此视图是一个Tableau可视化产品最基本的组成单元。

本节主要介绍在工作表里如何创建单个视图，在作图之前我们先认识Tableau创建视图的功能区和视图区，见图2-8。红色范围部分是创建视图的主要功能区，其中左边是卡功能区，从上至下依次为页面卡、筛选器卡和标记卡，标记卡包含了许多小的按钮如颜色、大小、标签等；上方红色框部分为行列功能区，将数据窗口中的字段拖放到此处就会在视图区显示相应的轴或标题。黑色框部分就是视图区，当我们使用卡和行列功能区进行操作时，图形的变化都会显示在视图区。

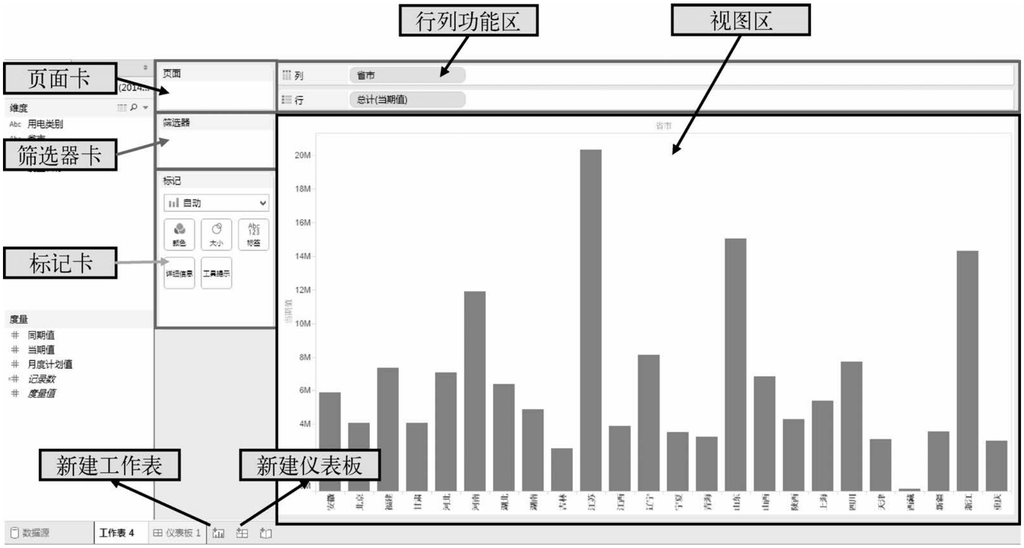


图2-8　认识视图工作区（另见彩插图2-8）

说明　视图中的图形单元我们称之为标记，比如圆图的一个圆点或柱形图的一根柱子，都是标记。

认识了视图功能区后，我们便可以利用数据窗口中的数据字段创建视图了。Tableau作图非常简单，拖放相关字段到相应的功能区，Tableau就会自动依据功能区相关功能将图形即时显示在视图区中。

2.3.1　行列功能区

我们以制作各省当期售电量柱形图为例，选定字段“省市”，用鼠标左键按住后拖放到列功能区，这时横轴就按照各省名称进行了分区，各省市成为了区标题，见图2-9。

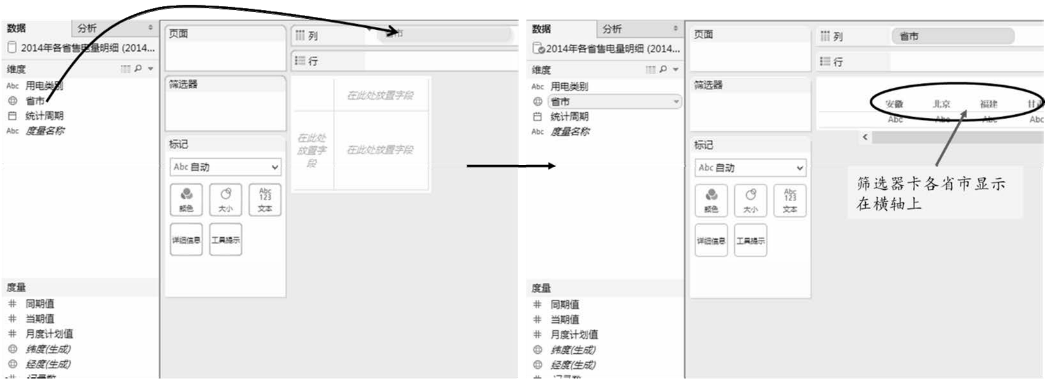


图2-9　拖放省市字段到列功能区

同理，拖放字段“当期值”到行功能区，这时字段会自动显示成“总计（当期值）”，视图区显示的便是售电量各省累计值柱形图，见图2-10。



图2-10 拖放当期值字段到行功能区

当然行列功能区可以不止拖放一个字段，例如我们可以将字段“同期值”拖放到“总计（当期值）”的右边，Tableau这时会根据度量字段“当期值”和“同期值”分别作出对应的轴，结果如图2-11所示。

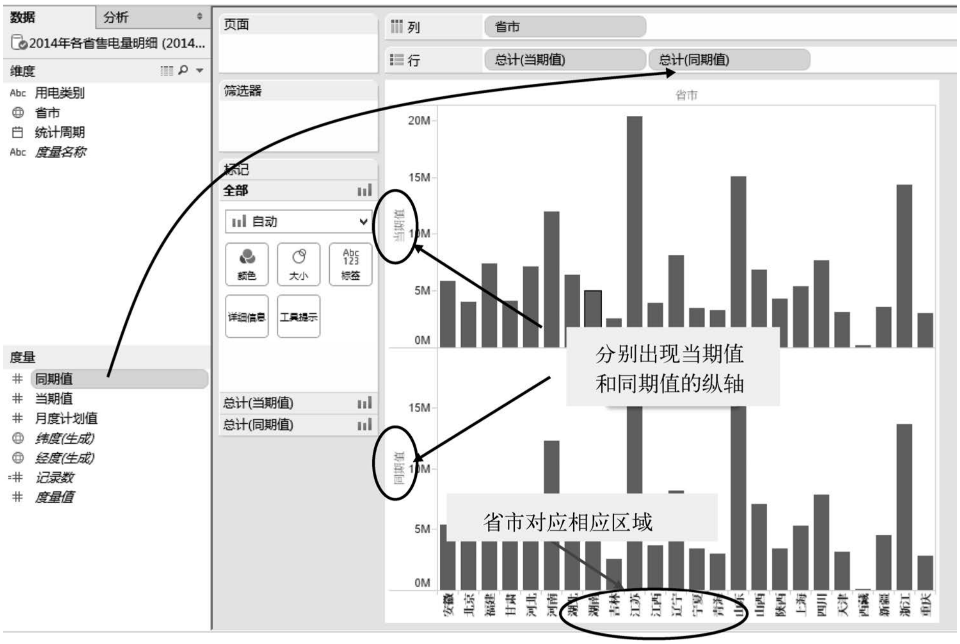


图2-11 在行功能区添加“同期值”字段

维度和度量都可以拖放到行功能区或列功能区，只是横轴、纵轴的显示信息会相应地改变，比如对于图2-11，我们可以单击工具栏上的 ，将行、列上的字段互换，这时省市显示在纵轴，

横轴变成了当期值和同期值，如图2-12所示。

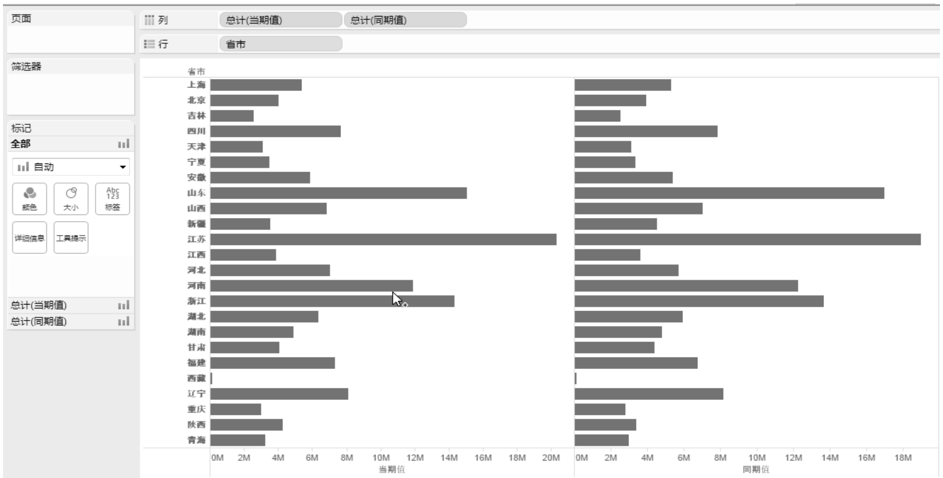


图2-12　互换行列字段

拖放度量字段“当期值”到功能区，字段会自动显示成“总计（当期值）”，这反映了Tableau对度量字段进行了聚合运算，缺省的聚合运算为总计。Tableau支持多种不同的聚合运算，如总计、平均值、中位数、最大值、计数等。如果想改变聚合运算的类型，比如想计算各省的平均值，只需在行功能区或列功能区的度量字段上，右键“总计（当期值）”或单击右侧小三角形，在弹出对话框中选择“度量”➤“平均值”即可，见图2-13。

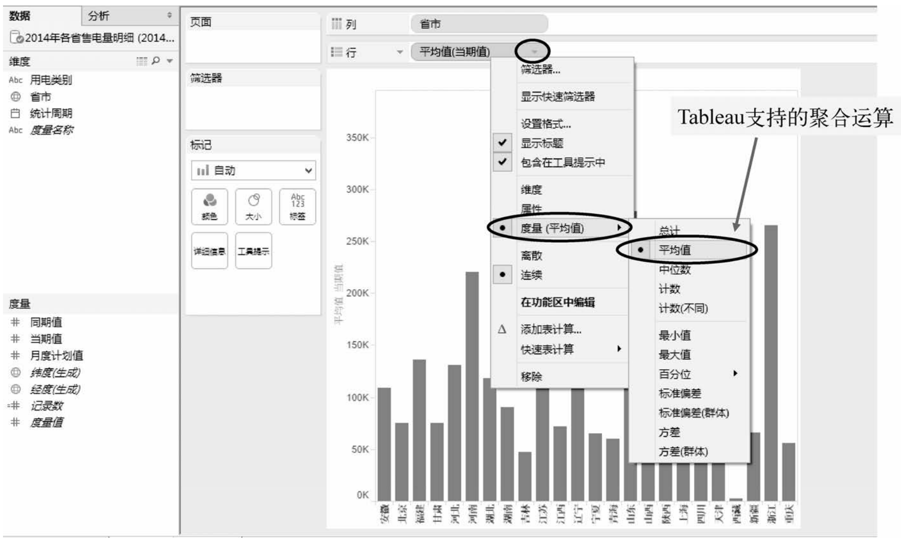


图2-13　度量字段的聚合运算

说明　Tableau求平均值是对行数的平均，以上海为例，其平均值为当期值总和除以省市为上海的行数，在原数据中每个省有6个月的当期值，每个月又分为9个用电类别，则出现上海的总行数为6×9=54，即平均值=总计/54。

2.3.2　标记卡

创建视图时，经常需要定义形状、颜色、大小、标签等图形属性。在Tableau里，这些过程都将通过操作标记卡来完成。标记卡样式见图2-14，其上部为标记类型，用以定义图形的形状。Tableau提供了多种类型的图以供选择，缺省状态下为条形图。标记类型下方有5个像按钮一样的图标，分别为“颜色”“大小”“标签”“详细信息”和“工具提示”。这些按钮的使用非常简单，只需把相关的字段拖放到按钮中即可，同时单击按钮还可以对细节、方式、格式等进行调整。此外还有3个特殊按钮，特殊按钮只有在选择了对应的标记类型时，才会显示出来。这3个特殊按钮分别是线图对应的路径、形状图形对应的形状、饼图对应的角度，见图2-15。



图2-14　标记卡和标记类型



图2-15　特殊标记按钮

1. 颜色、大小和标签

拖放“省市”到列功能区，拖放“当期值”到行功能区，完成最简单的显示各省售电量累计值的柱形图。这时如果想让不同省市显示不同颜色，可利用标记卡中的颜色来完成，只需将字段“省市”拖放到颜色里即可（见图2-16）。

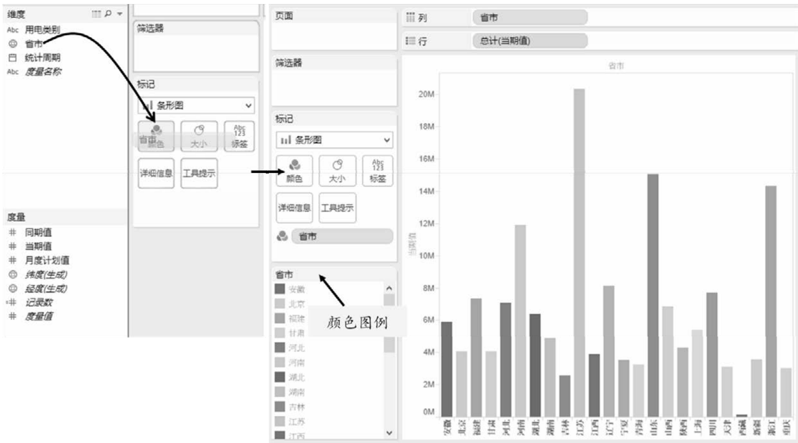


图2-16　颜色图例（另见彩插图2-16）

这时，卡功能区的下方会自动出现颜色图例，用以说明颜色与省市的对应关系。单击颜色图例右上角处，在弹出框中可以对颜色图例进行设置，如编辑标题、排序、设置格式等。其中单击选项“编辑颜色”，进入颜色编辑页面，可以对不同的区域自定义不同的颜色。比如要将上海的蓝色改为红色，可选择“编辑颜色”进入颜色编辑页面。首先左键单击上海，然后左键单击右侧选择调色板的红色，确定即可，如图2-17所示。

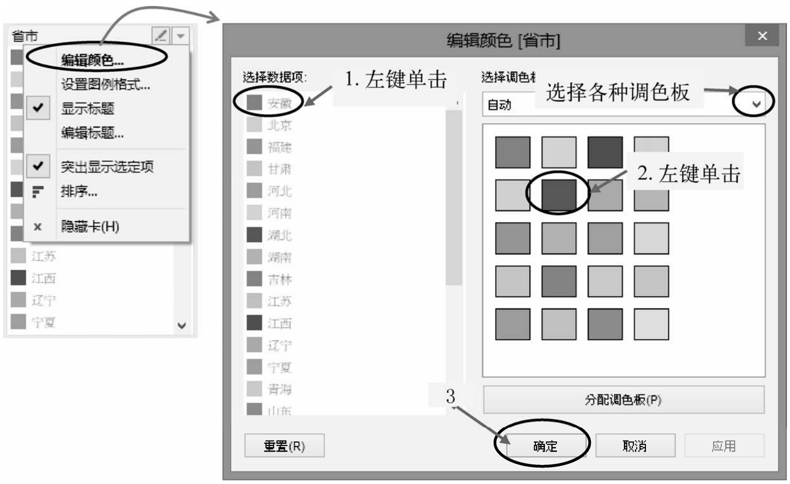


图2-17　编辑颜色

如果要对视图中的标记添加标签，如将当期值添加为标签显示在图上，只需将字段“当期值”拖放到标签即可，如图2-18所示。

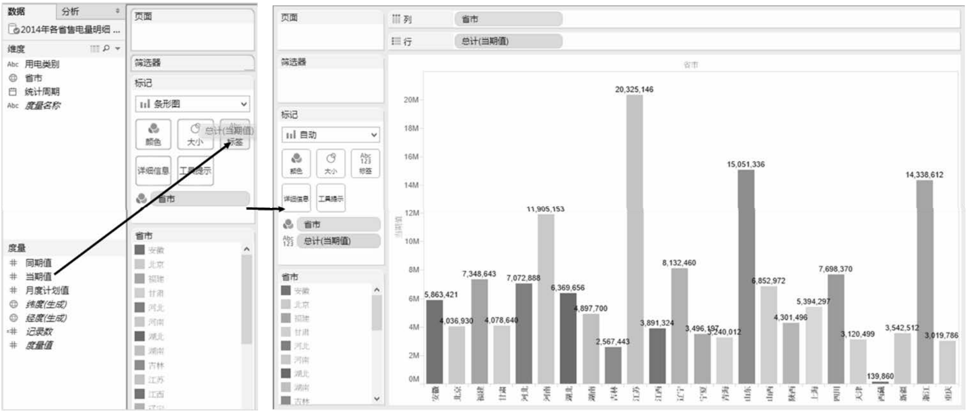


图2-18 添加标签

标签显示的是各省的当期值总计，如果想让标签显示各省当期值的总额百分比，可右键单击“标记”卡中的总计（当期值）或单击总计（当期值）右侧小三角标记，在弹出的对话框中选择“快速表计算”➤“总额百分比”，这时视图中的标签将变为总额百分占比，如图2-19所示。此外，单击标签，可对标签的格式、表达方式等进行设置。

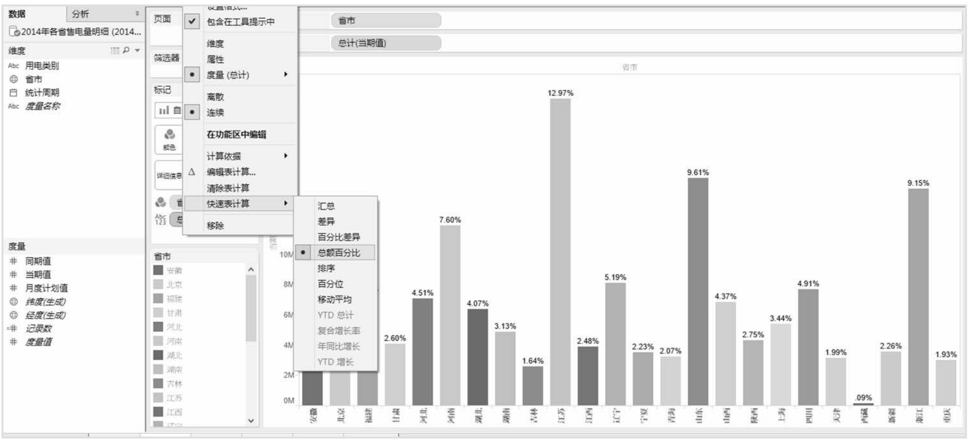


图2-19 标签显示为总额百分比

大小和颜色类似，拖放字段到“大小”，视图中的标记会根据该字段改变大小，这里不再详细阐述。需要注意的是，颜色和大小只能放一个字段，但是标签可以放多个字段。

2. 详细信息

详细信息可能比较难理解，它的功能是依据拖放的字段对视图进行分解细化。我们以圆图为

案例，将“省市”拖放到列功能区，“当期值”拖放到行功能区，标记类型选择“圆”图，如图2-20所示。这时每个圆点所代表的值其实是9个用电类别6个月的总和。

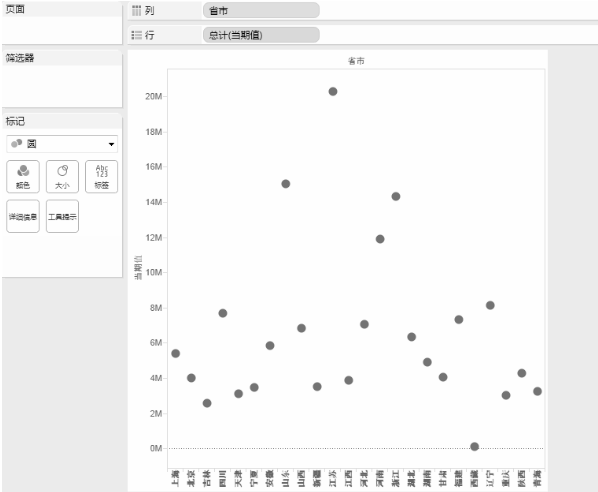


图2-20　详细信息

拖放字段“用电类别”到“详细信息”，Tableau会依据“用电类别”进行分解细化，这时每个圆点变为9个圆点，每一个点代表相应省市某一用电类别的总和，如图2-21所示。这时候每个点是一个用电类别6个月的总和，我们拖放字段“统计周期”到“详细信息”并选择按“月”（Tableau默认的是按“年”），这时每个点再次解聚为6个点，每个点表示该省某月某用电类别总和，如图2-22所示。

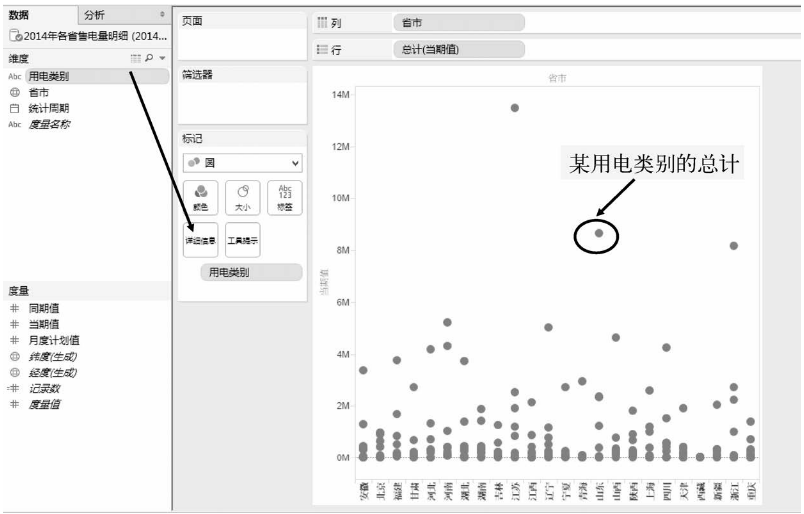


图2-21　依据“用电类别”的详细信息

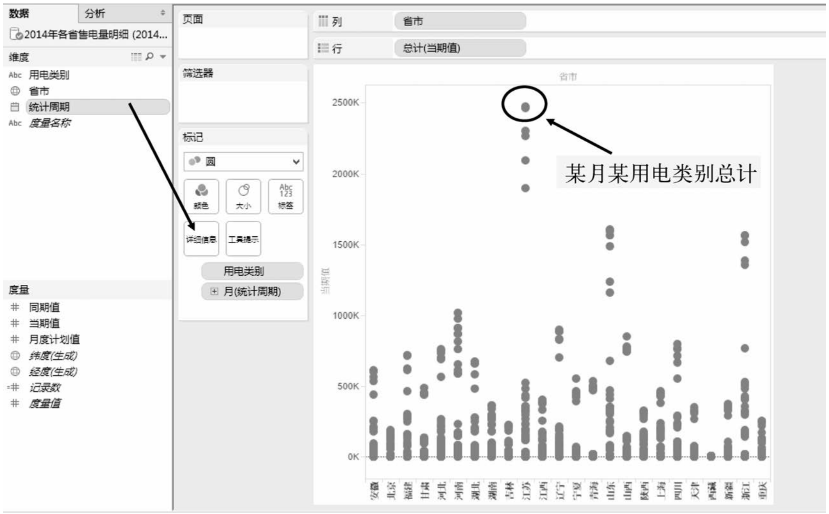


图2-22　依据“用电类别”和“月（统计周期）”的详细信息

其实直接拖放到“标记”卡的下方就可以表示详细信息，并且颜色、大小、标签都具有与详细信息搭配使用的功能。以颜色为例，在图2-22的基础上拖放字段“用电类别”到“颜色”，视图中每一个点也会解聚为9个颜色不同的点，如图2-23所示。

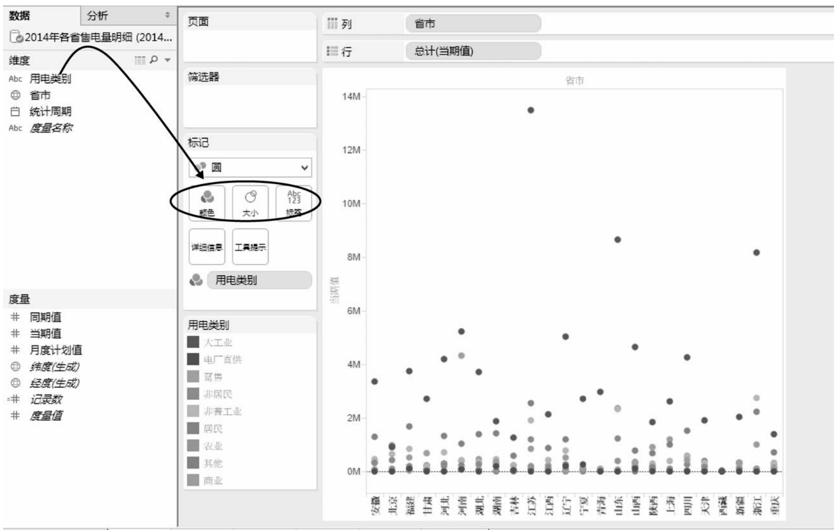


图2-23　依据颜色的详细信息（另见彩插图2-23）

3. 工具提示

当鼠标移至视图中的标记上时，会自动跳出一个显示该标记信息的框，以图2-23为例，当将

鼠标移至视图中某个圆点时，边上就会自动出现提示信息（见图2-24），这便是工具提示的作用。

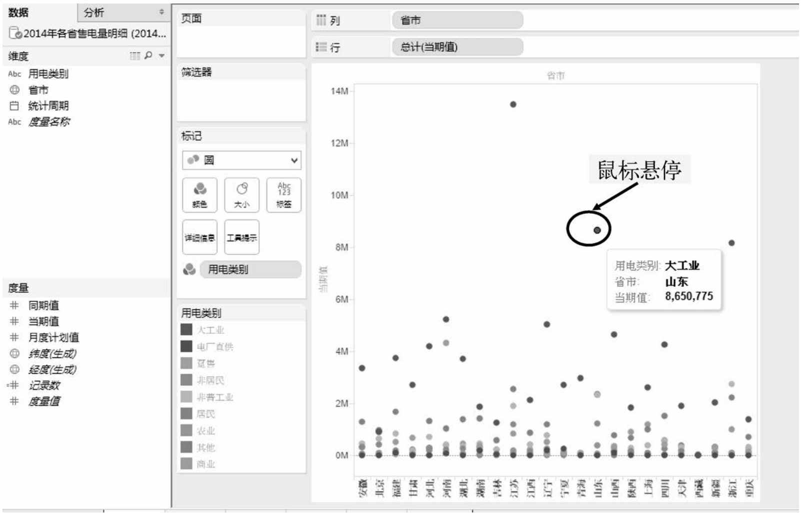


图2-24　工具提示

左键单击“工具提示”可以看到工具提示的内容，可对这些内容进行删除、更改格式、排版等操作，见图2-25（左）。Tableau会自动将“标记”卡和行列功能区的字段添加到工具提示中，如果还需要添加其他信息，只需将相应的字段拖放到“标记”卡中。例如我们将字段“同期值”和“月度计划值”拖放到“标记”卡下方，这时单击“工具提示”，就可以看到同期值和月度计划值的信息已经出现在里面了，见图2-25（右）。



图2-25　添加工具提示

说明 本节主要是对Tableau创建视图需要的一些共性功能键如颜色、大小、标签等进行介绍，而对于标记卡中的特殊按钮如路径、角度、形状的使用，详见后续章节。其中路径的介绍见5.3.1节，角度的介绍见4.3节。

2.3.3 筛选器

有时候只想让Tableau展示数据的某一部分，如只看1月份的售电量、只看华北和华中地区各省情况、只用电量大于700 000的数据等，这时可通过筛选器完成上述选择。拖放任一字段（无论维度还是度量）到筛选器卡里，都会成为该视图的筛选器。以图2-25为例，如果让视图里只显示大工业的点，或者只显示居民和非居民的点，只需要将字段“用电类别”拖放到筛选器卡里，这时Tableau会自动弹出一个对话框，单击“从列表中选择”选项就会显示“用电类别”的内容，这里可直接勾选想展现的用电类别，如大工业。单击“确定”后字段“用电类别”就显示在筛选器中了，见图2-26。



图2-26 添加筛选器

将字段拖放到筛选器卡里后，需要将筛选器显示出来，右键或单击右侧小三角形，在对话框中选择“显示快速筛选器”即可。这时工作表的右边会显示筛选器，即可进行筛选操作。例如图2-27只显示了大工业类别电量信息，单击筛选器右上角的小三角形可以对筛选器进行设置，包括筛选器的表现形式、功能选项等。这里介绍两个常用的选项（图中右边框出部分）：①应用于工作表，可以选择将该筛选器应用于多个表（前提是这几个工作表有共同的筛选器），这个功能在

仪表板中经常使用，这里不过多介绍，详见9.2节的步骤16；②筛选器的表现形式，Tableau提供了单值（列表）（每次只能选择一个值）、单值（下拉列表）、多值（列表）（同时可以选多值）等选项，默认是多值（列表）。

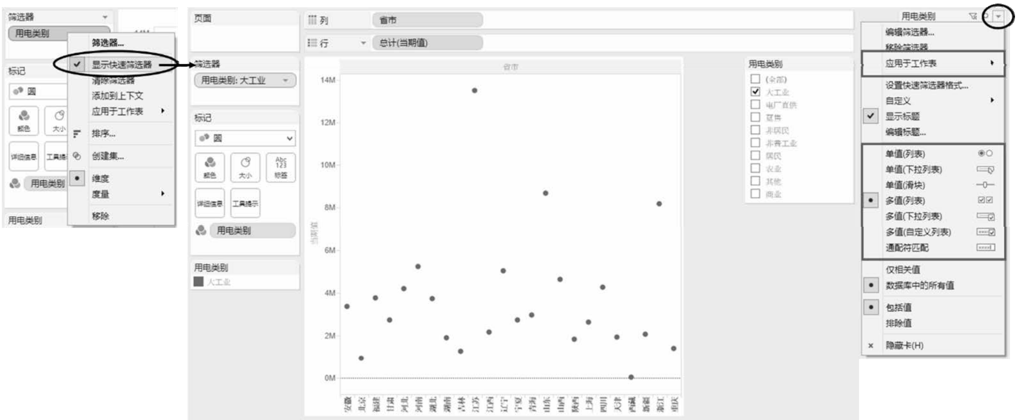


图2-27　显示快速筛选器

Tableau提供了多种筛选方式，在图2-28筛选器弹出框上方可以看到 “常规”“通配符”“条件”和“顶部”选项，每一个选项之下都有相应的筛选方式，这大大丰富了筛选操作形式。比如在9个用电类别中想看当期值总计在前3位的用电类别，单击“顶部”，选择“按字段”➤“顶部”➤3、以及“当期值”➤“总计”，见图2-28。根据以上操作，图形显示用电量当期值居前3位的用电类别分别是大工业、居民和非普工业。

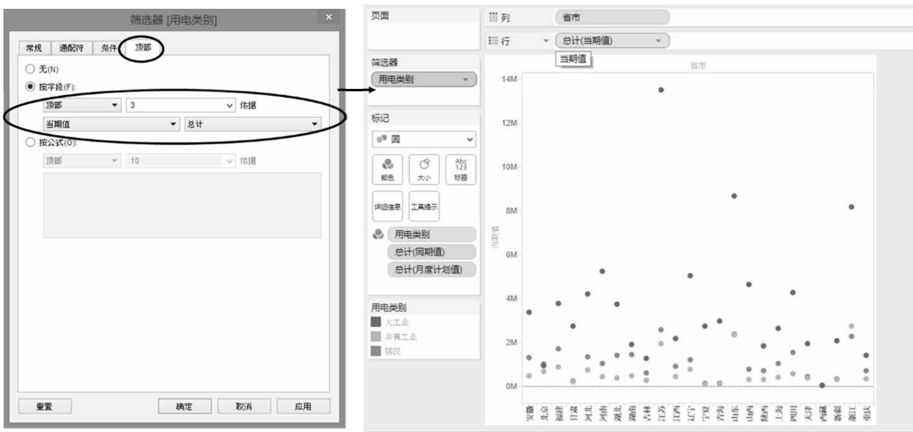


图2-28　筛选方式

将度量设置为筛选器的操作方法与此相同，只是这时候筛选依据的是连续的值。比如要显示售电量在700 000到10 000 000之间的点，见图2-29。

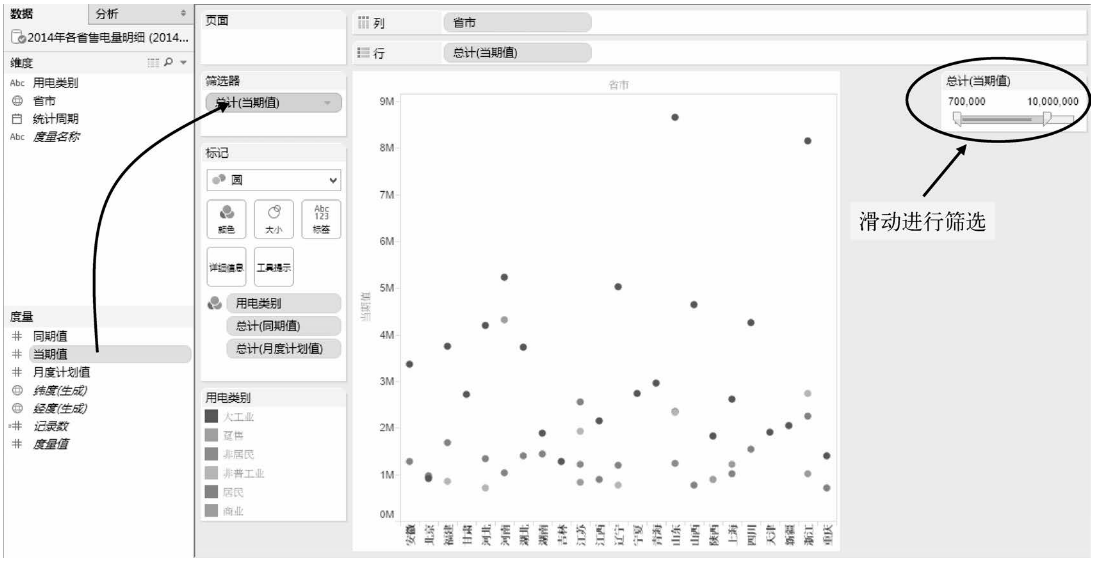


图2-29 显示度量筛选器

说明 对于筛选器的使用，其实有更简单的方法：直接将鼠标移至数据窗口中需要用作筛选器的字段，右键选择“显示快速筛选器”即可。

2.3.4 页面

将一个字段拖放到页面卡会形成一个页面播放器，播放器可让工作表更灵活。

为了更好地展示页面功能，我们新建一个工作表，拖放字段“统计周期”到列，Tableau会默认“统计周期”为年，我们要手动转换为月，拖放“当期值”到行，标记类型选择为圆，见图2-30。

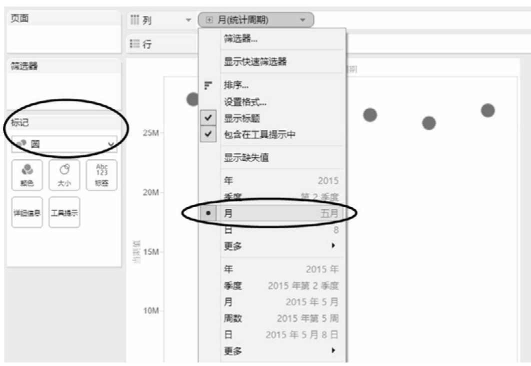


图2-30 页面播放器

拖放字段“统计周期”到页面卡，这时页面卡下方会自动出现一个“年（统计周期）”的播放器。将日期的显示“年（统计周期）”调整为“月（统计周期）”，见图2-31。

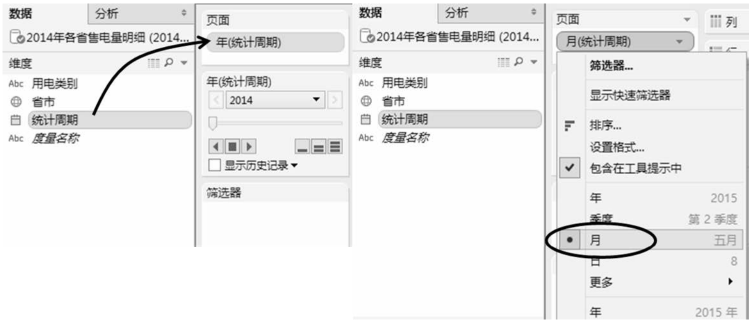


图2-31　调整页面播放器

单击播放器的播放键，可以让视图动态播放出来，选择“显示历史记录”可以设置播放的效果，见图2-32。



图2-32　播放器演示

2.3.5　智能显示

在Tableau的右端有一个智能显示的按钮，单击展开，如图2-33所示，显示了24种可以快速创建的基本图形。将鼠标移动到任意图形上，下方都会显示做该图需要的字段要求，如将鼠标移动

到符号地图上，下方会显示“1个地理维度，0个或多个维度，0至2个度量”，这表明创建该视图必须要一个地理类型的字段类型，度量不能超过2个。

按照要求我们将地理维度“省市”和字段“当期值”拖放到行列功能区，这时候发现智能显示的某些图形高亮了，高亮的图形表示用目前的字段可以快速创建的图形，见图2-34。单击符号地图，这时符号地图就创建完成了。这时候发现行列功能区变为经纬度字段，“省市”在“标记”卡中表示详细信息，符号大小表示“当期值”，见图2-35。



图2-33　智能显示

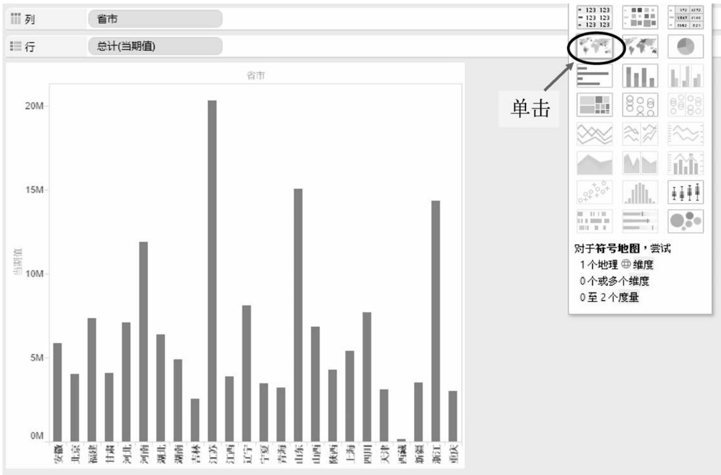


图2-34　快速创建图形



图2-35　快速创建符号地图

2.3.6　度量名称和度量值

度量名称和度量值都是成对使用的，目的是将处于不同列的数据用一个轴展示出来。当想同

时看各省当期值和同期值时，需按照2.3.1节所述，拖放“省市”到列功能区，再分别拖放“同期值”和“当期值”到行功能区，图形显示见图2-36。可以看到，图中出现了当期值和同期值两条纵轴。

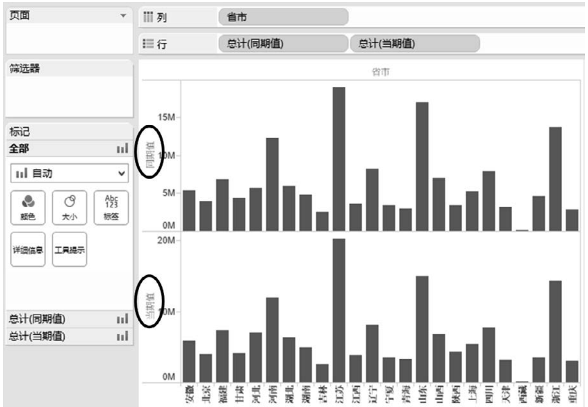


图2-36　显示两条纵轴

下面我们利用度量值和度量名称来完成两列不同数据共用一个轴的操作。首先还是拖放字段“省市”到列功能区，然后拖放度量值到行功能区，这时在左下方会显示度量值包含了哪些度量，Tableau默认的度量值会包含所有的度量。由于我们只需要当期值和同期值，因此右键行上的度量值或单击度量值右边的小三角形，选择“筛选器”，去掉记录数前面的钩，只保留当期值和同期值，操作见图2-37。

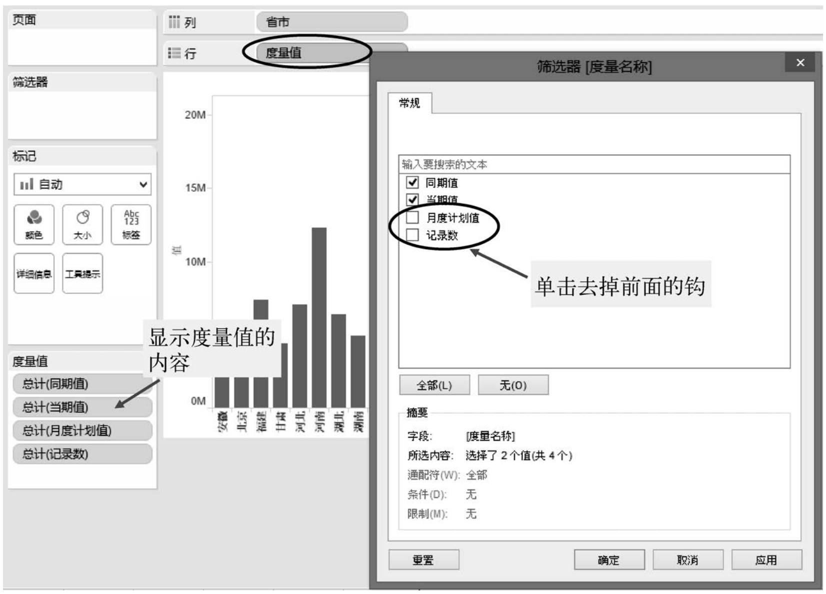


图2-37　筛选度量值

这时可以看到度量值里只包含了当期值和同期值。将度量名称拖放到“颜色”，这时柱状图按颜色分成了当期值和同期值，二者共同一个纵轴。如果习惯将当期值和同期值分开为两个柱子，只需将度量名称拖放到列功能区，放置在省市的右边，见图2-38。

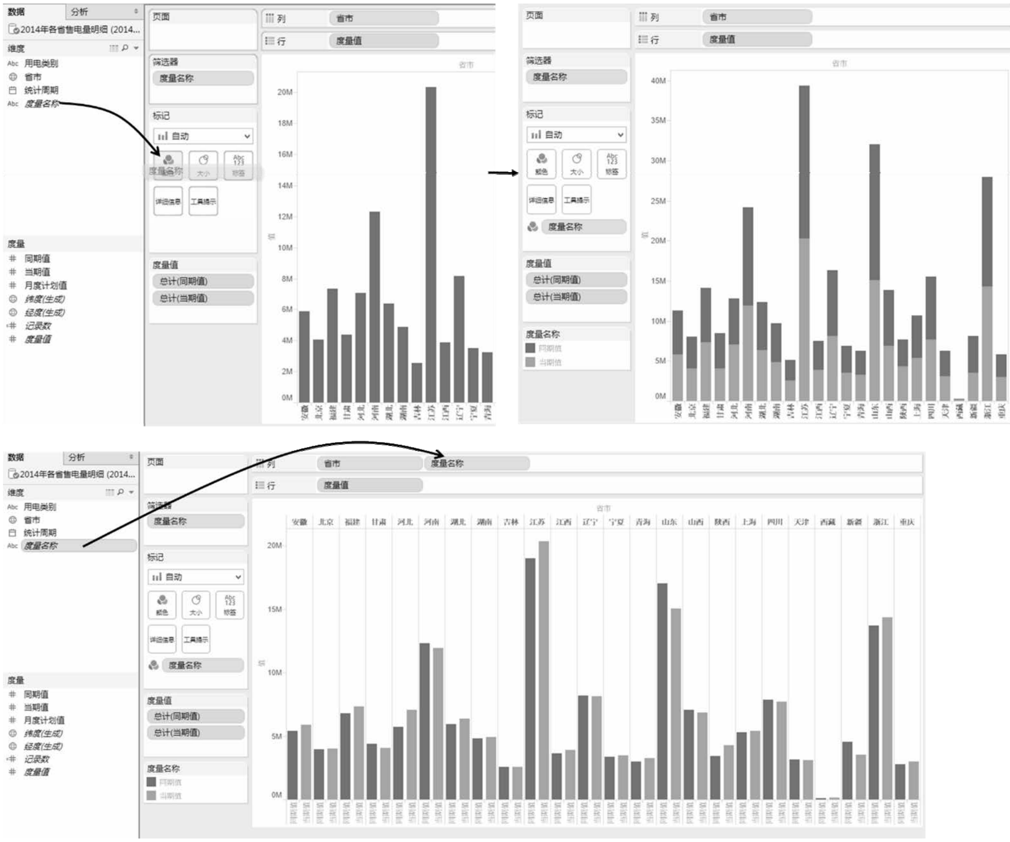


图2-38　度量值按颜色显示成双柱图

说明　事实上，我们可以利用智能显示快速完成双柱图形，在智能显示里双柱图称为并排图，把鼠标放上去会显示完成该图需要“1个或多个维度，1个或多个度量，至少需要3个字段”。我们将“省市”拖放到列功能区，将“当期值”和“同期值”拖放到行功能区，这时并排图被高亮，单击即可完成图2-38所示的图形。

2.4　创建仪表板

完成所有工作表的视图后，我们便可以将其组织在仪表板中了。单击下方的新建仪表板，进入到仪表板工作区。创建仪表板也是用拖放的方法，将创建好的工作表拖放到右侧排版区（见

图2-39），并按照一定的布局排版好，最后添加操作完成互动设置。以图2-37、图2-38和图2-39为例完成的一个简单仪表板，见图2-40。创建仪表板的详细操作请参见第9章。

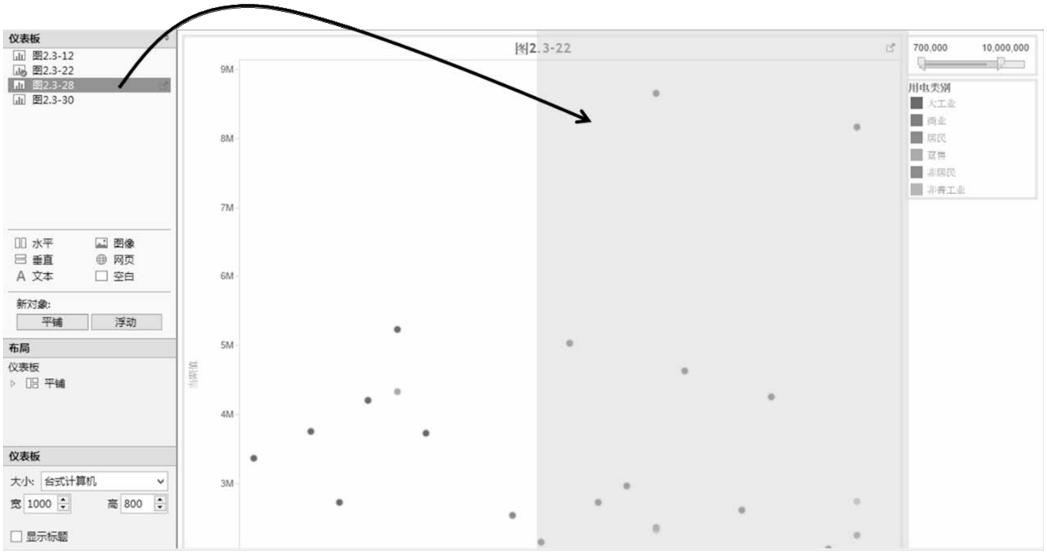


图2-39　拖放工作表到仪表板

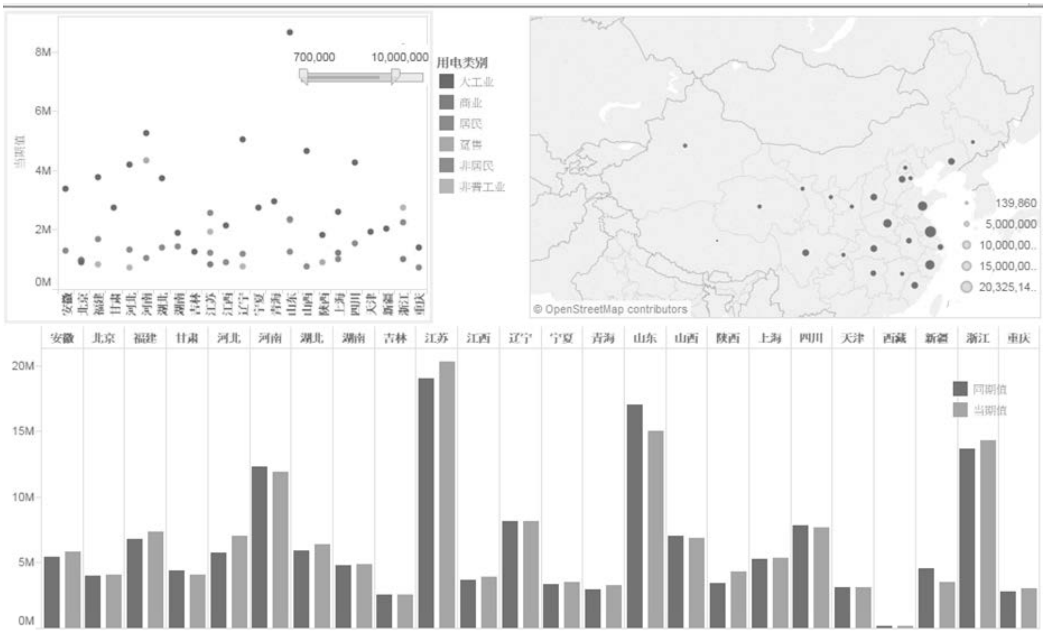


图2-40　创建简单仪表板

2.5 保存工作成果

创建完仪表板后，应当将结果保存在Tableau工作簿中。如图2-41所示，选择“文件”➤“保存”或快捷键Ctrl+S，进行保存。保存的类型可以是Tableau工作簿（*.twb），该类型将所有工作表及其连接信息保存在工作簿文件中但不包括数据；也可以是Tableau打包工作簿（*.twbx），该类型包含所有工作表、其连接信息以及任何其他资源如数据、背景图片等。



图2-41 保存工作成果

本章以一个简单案例介绍了Tableau从连接数据到最后工作簿发布的过程，重点介绍了如何利用功能区创建视图，以便读者熟悉Tableau拖放的作图方法。Tableau详细的使用方法，可从第3章开始逐步学习。Tableau支持多种数据源，第3章介绍了详细的数据连接情况；Tableau可以创建丰富的视图，具体内容可以参考第4章初级可视化分析、第5章地图分析、第6章高级数据操作以及第7章利用Tableau的高级特性进行高级可视化分析；详细的仪表板创建可参考第9章分析图表整合；第10章重点介绍了Tableau分析成果共享。

第 3 章

数据连接与管理

3

连接数据源是利用Tableau进行数据分析的第一步，Tableau拥有强大的数据连接能力，支持几乎所有的主流数据源类型。本章将从连接最简单的电子表格数据开始，重点说明如何通过Tableau快速连接到各类数据源，如何实现多表联接查询和多数据源数据关联，如何创建和管理数据提取，以及如何管理和操作数据源。

3.1 Tableau 的数据架构

传统BI软件的元数据设计要么一次性建立完整企业级的元数据体系，要么完全不进行元数据管理。Tableau的元数据管理更加灵活，可以细分为数据连接层（Connection）、数据模型层（Data Model）和数据可视化层（VizQL）。其中，可视化层中使用的VizQL是以数据连接层和数据模型层为基础的Tableau核心技术，对数据源（包括数据连接层和数据模型层）非常敏感。

Tableau这样的三层设计，既可以让不了解元数据管理的普通业务人员进行快速分析，又方便了专业技术人员进行一定程度的扩展。为了更好地理解Tableau处理数据的方式与能力，下面我们从数据新建数据源模型开始来理解Tableau中的数据。

1. Tableau中的数据连接层

数据连接层决定了如何访问源数据和获取哪些数据。数据连接层的数据连接信息包括数据库、数据表、数据视图、数据列，以及用于获取数据的表连接和SQL脚本，但是数据连接层不保存任何源数据。

在数据连接层，用户可以方便地对Tableau工作簿的数据连接进行修改，例如，将一系列仪表板的数据连接从测试数据库切换到生产数据库，只需要编辑数据连接，变更连接信息，Tableau会自动处理所有字段的实现细节。

在数据连接层中，不论是在Tableau Server还是在Tableau Desktop，支持的数据类型都非常丰富。

Tableau支持传统的关系数据源（MySQL、Oracle、IBM DB2等）、多维数据源（Oracle Essbase、Microsoft Analysis Services、Teradata OLAP Connector）、Hadoop系列产品中的数据源（Cloudera Hadoop、Hortonworks Hadoop Hive、MapR Hadoop Hive等）、Tableau数据提取、Web数据源（Google

Analysis、Google BigQuery、Salesforce等）、常见的本地文件（Excel、文本文件等）等多种类别，如图3-1所示。可通过Tableau Desktop直接新建数据源，也可以通过Tableau Server新建数据源，还可以把数据源发布到Tableau Server。

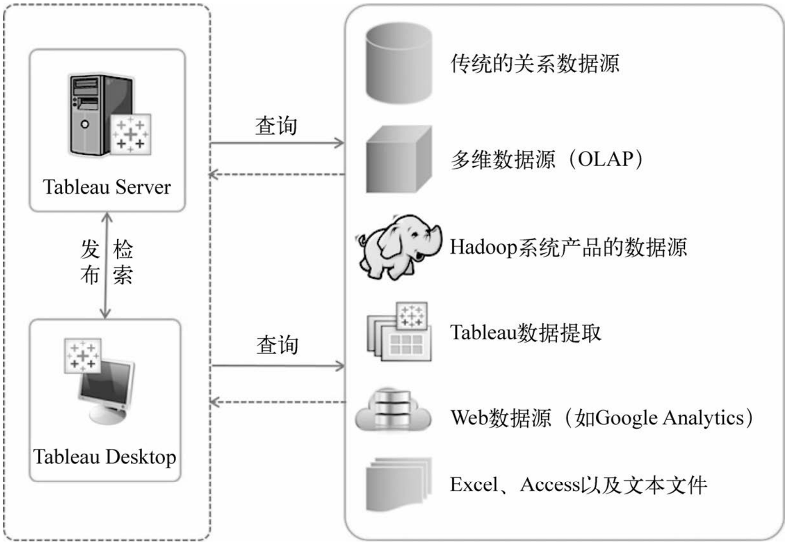


图3-1 数据连接层中的数据源

2. Tableau中的数据模型层

不论数据源来自哪种服务器，Tableau中的数据都会分为维度和度量两大类。多维数据源的使用者对这两个概念应该非常熟悉，关系数据库中的数据可以在Tableau中进行一定程度的数据建模工作。

这些数据建模工作可以在Tableau的数据模型层完成，主要内容包括管理字段的数据类型、角色、默认值、别名，以及用户定义的计算字段、集和组等。例如，如果在数据库中删除字段，那么在Tableau工作表中对应的字段会被自动移除，或者自动映射到别的替代字段。

在完成数据连接后，Tableau会自动判断字段的角色，把字段分为维度字段和度量字段两类。如果所连接数据是多维数据源，那么Tableau直接获取数据立方体维度和度量信息；如果连接的是关系数据源，Tableau会根据数据库的数据来判断该字段是维度字段还是度量字段。

为了更好地说明Tableau数据模型中的维度字段与度量字段，我们可以查看一个多维数据源的案例，如图3-2所示。

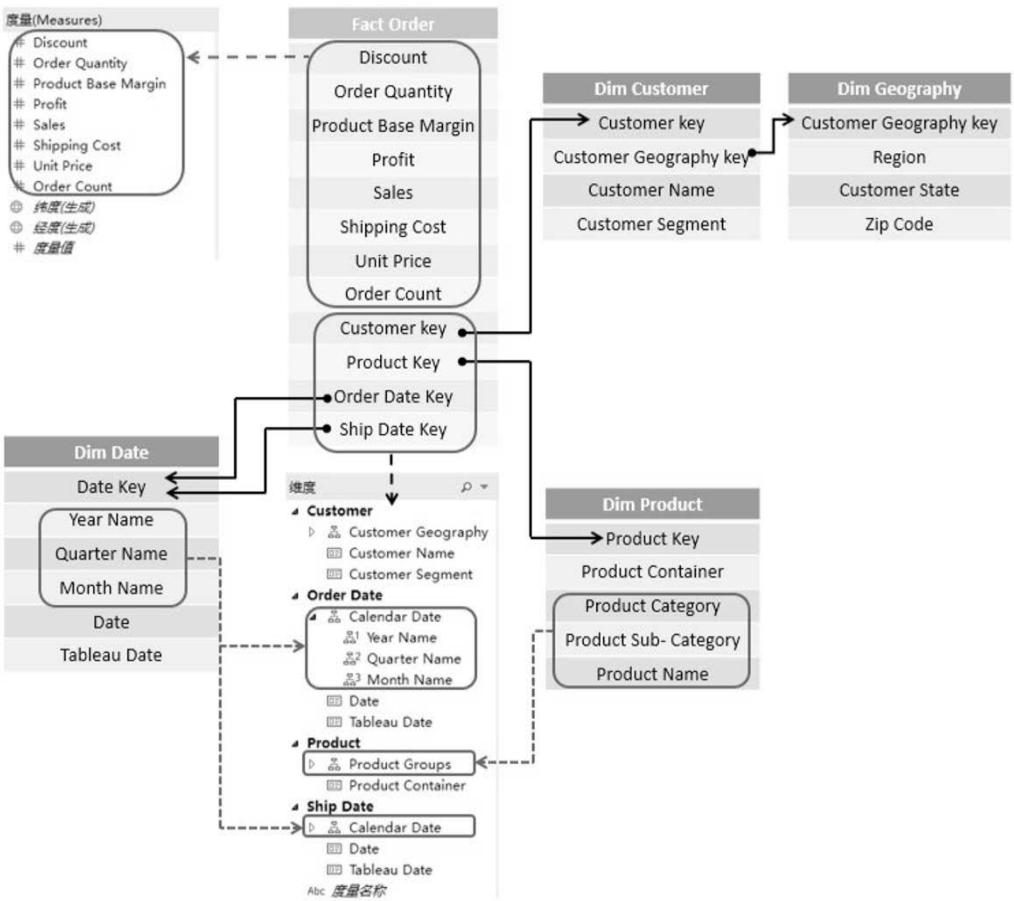


图3-2　多维数据源案例

在一般的数据建模过程中，首先要对业务数据库进行ETL操作（即数据抽取、转换和加载），采用的结构通常为星形结构或者雪花形结构，然后定义维度中的层级关系和属性，形成数据立方体（Cube）。这里事实表（Order）中的4个属性键在Tableau中分别对应Tableau里的4个维度（Customer、Product、Order Date和Ship Date），而Tableau中的度量取自事实表中的另外8个字段。

Tableau可以识别出多维数据源中预先定义好的分层结构。在维度表（Date、Product、Geography）中，可以事先定义层级关系和属性，例如可以将Date表中的Year Name、Quarter Name、Month Name制作成Calendar Date分层结构。

值得注意的是，由于多维数据源的特性，Tableau引入的多维数据源本身已经是一种聚合的形式，无法再进行进一步的聚合，并且维度字段将不能随意改变组织形式（如分组、创建分层结构、角色转换）和参与计算，同时度量字段也不能使用分级和改变角色。

3.2 数据连接

要在Tableau中创建视图，首先需要新建数据源，如图3-3所示。

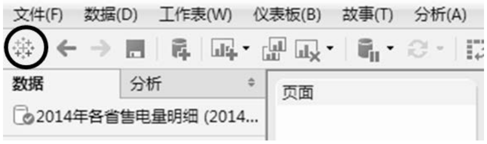


图3-3 新建数据源

打开Tableau软件，进入主界面之后在页面左侧选择“新建数据源”，也可以在主界面菜单栏选择“数据”➤“新建数据源”，在下级界面的左侧会看到Tableau支持的数据源类型。由于数据类型较多，图3-4仅展示了一部分文件和服务器数据源。



图3-4 部分数据源类型

3.2.1 连接文件数据源

本节主要介绍如何通过Tableau快速连接到电子表格、Access、Tableau工作簿等各类文件数据源。

1. 连接到电子表格

在文件数据源中，最常用的是电子表格，下面以Microsoft Excel文件为例进行说明。选择Microsoft Excel，在新的界面中可以看到数据源内含有两张Excel工作表，在弹出的“打开”对话框中选择“Transformer Load info.xls”，单击“打开”，如图3-5所示。

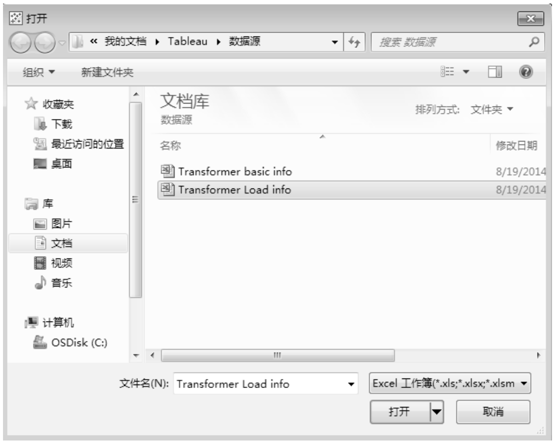


图3-5　连接Transformer Load info示例

根据界面上部“将工作表拖到此处”的文字提示，如图3-6所示，将表“重过载信息”拖入界面中部框内（双击此表也可达到相同的效果），这时可在界面下方看到“重过载信息”工作表的数据，见图3-7。



图3-6　“编辑数据源”界面

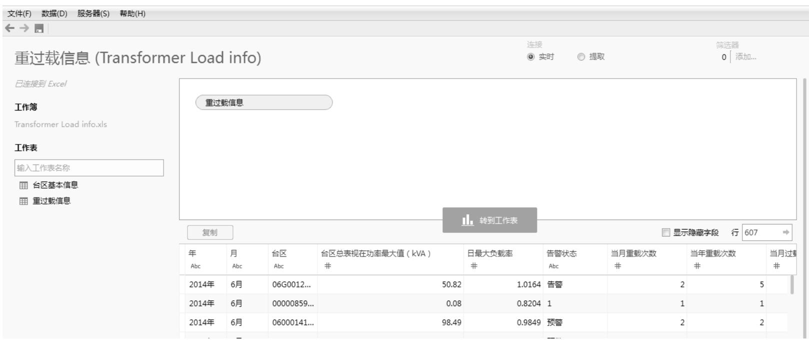


图3-7　选择工作表

在确认表中数据信息无误后选择“新建数据表”，随即进入工作区界面，如图3-8所示，此时即成功连接到了Excel数据源。

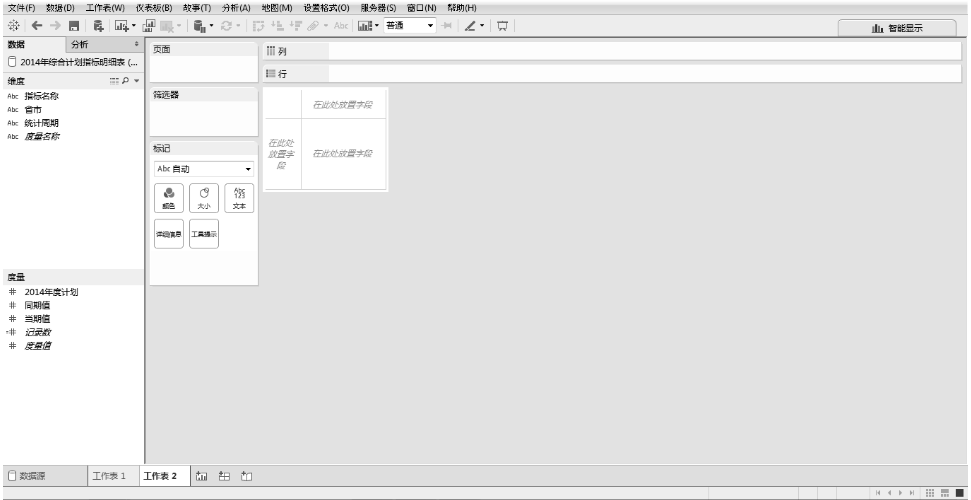


图3-8 工作区界面

如果需要在下次使用时快速打开数据连接，可以将数据连接添加到已保存数据源中，操作步骤为选择“数据” ➤ “<数据源名称>” ➤ “添加到已保存的数据源”，在弹出的窗口中选择“保存”，如图3-9所示。



图3-9 添加到已保存的数据源

再次打开Tableau之后，在开始界面就可以直接连接到已保存数据源，如图3-10所示。

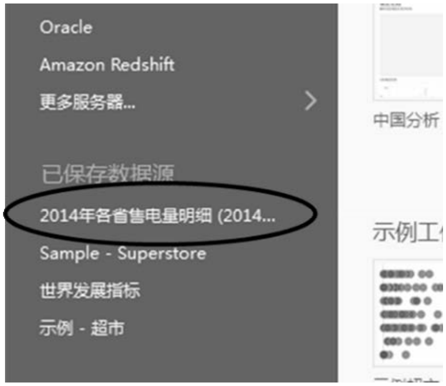


图3-10　添加在“已保存数据源”列表

2. 连接到Access文件

连接到Microsoft Access数据源的操作步骤和连接到电子表格基本类似，且均在新建数据源界面进行实现。

和连接电子表格不同的是，在选定数据表的界面左下角会出现“新自定义SQL”选项，熟悉SQL的用户可以选择使用SQL查询连接数据，如图3-11所示。



图3-11　连接到Access

双击“新自定义SQL”[①]，在弹出的“编辑自定义SQL”文本框中键入或粘贴SQL查询语句。完成后单击“确定”，即可实现到Microsoft Access数据源的连接，如图3-12所示。

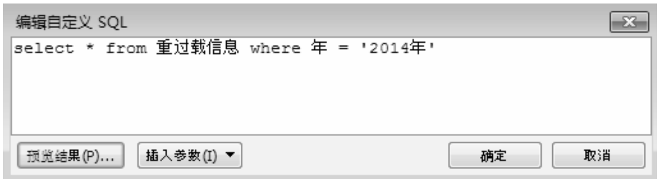


图3-12 编辑自定义SQL

3. 连接到Tableau工作簿

在新建数据源界面中单击“其他文件”，在弹出的对话框中打开工作簿或打包工作簿，Tableau会提示导入所选工作簿的数据源，单击“确定”完成到数据源的连接，如图3-13所示。

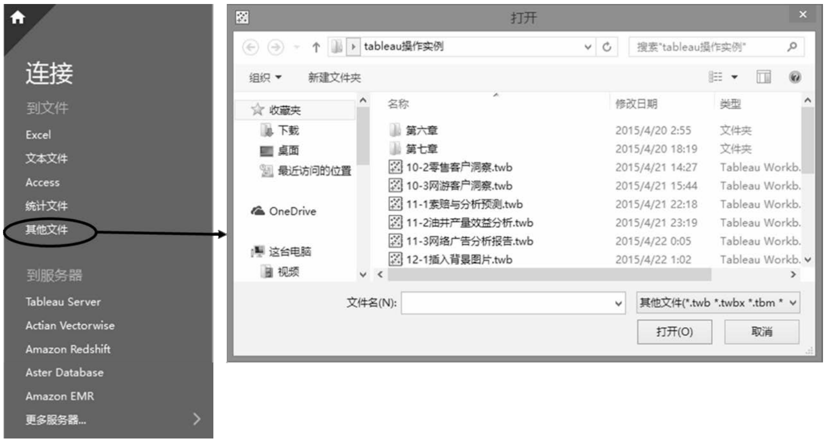


图3-13 导入所选工作簿的数据源

3.2.2 连接服务器数据源

在新建数据源界面中，标题“到服务器”下方列出了Tableau所支持的各类服务器数据源，用户可以根据需要进行选择。Tableau 9.1版本进一步支持对Web数据源及目前热门的几类云端数据库（如Amazon Aurora、Google Cloud SQL、Microsoft Azure）的连接。本节将选取Oracle、ODBC和Cloudera Hadoop为例加以说明。

1. 连接到 Oracle 数据库

选择位于“到服务器”标题下方的“Oracle”数据库，在界面右侧填写连接Oracle所需要的

① 新自定义SQL的相关应用在下文连接到ODBC中也会提到。

相关信息，完成后选择“连接”，如图3-14所示。



图3-14　连接到Oracle

还可以连接到Oracle Essbase，这是Oracle的多维数据库。与关系数据源不同，当连接到多维数据源时，不能使用SQL或MDX语言进行查询，而在Tableau创建实时可视化查询时会使用优化后的MDX语言进行操作。

2. 连接到 ODBC

选择位于“到服务器”标题下方的“其他数据库（ODBC）”，如图3-15所示。

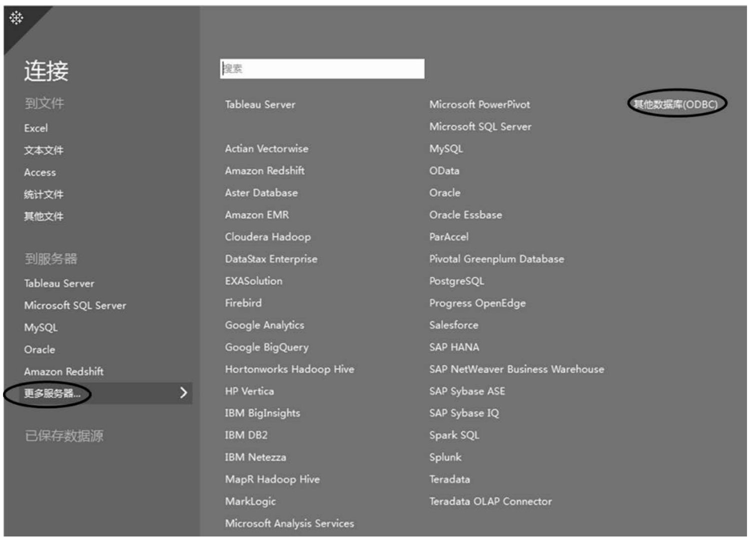


图3-15　连接到ODBC

在右侧选择驱动程序，在连接方式一栏完善信息，完成后单击“连接”，弹出“ODBC数据源配置”对话框。在对话框中完善数据源配置信息，如服务器地址及端口、账号、数据库类型等，确认无误后单击“OK”，如图3-16所示。



图3-16　ODBC连接信息窗口

单击“确定”后进入编辑数据源界面，输入数据库表的名称，拖入右侧窗口，选择“转到工作表”，随即进入工作区界面，完成数据连接。

和连接Microsoft Access一样，在Tableau中也可以通过新自定义SQL来完成数据连接，如图3-17所示。

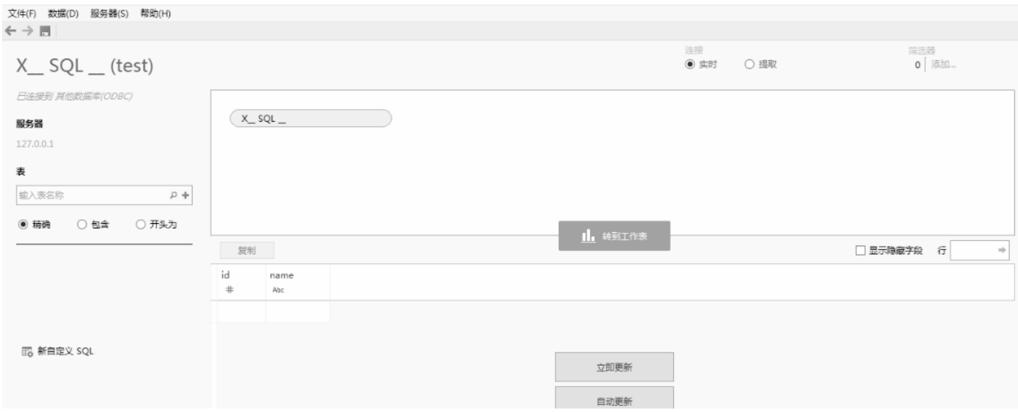


图3-17　通过自定义SQL查询新建数据源

3. 连接到 Cloudera Hadoop

选择位于“到服务器”标题下方的“Cloudera Hadoop”。在右侧输入服务器密码与端口号，并且在连接方式一栏完善信息，完成后单击“确定”（见图3-18）。

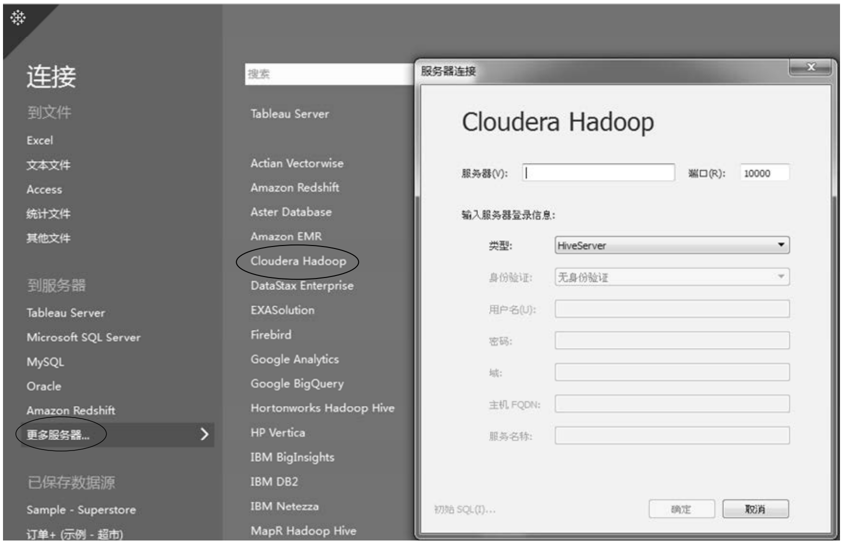


图3-18　登录Coudera Hadoop

需要说明的是，只要安装了相应驱动，就可以使用Hive和Impala两种连接方式（从查询效率考虑，建议读者使用Impala），如图3-19所示。

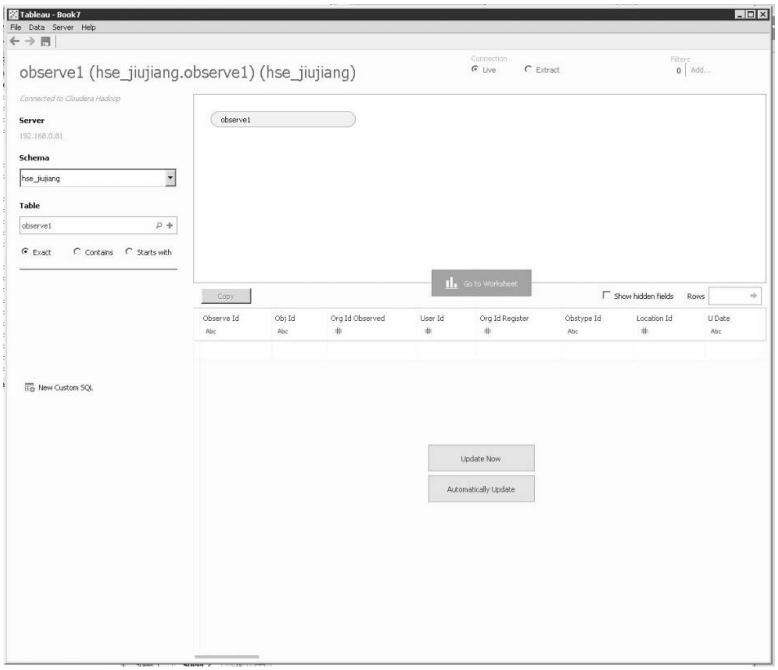


图3-19　选择数据源

然后，可以选择Hadoop中的表进行实时提取数据。虽然Hadoop不是传统意义上的SQL，但是借助Impala或Hive，也可以通过新自定义SQL来完成数据连接，并且从图3-20可以发现，Tableau从Hadoop中获得的数据和其他关系型数据源没有任何区别。

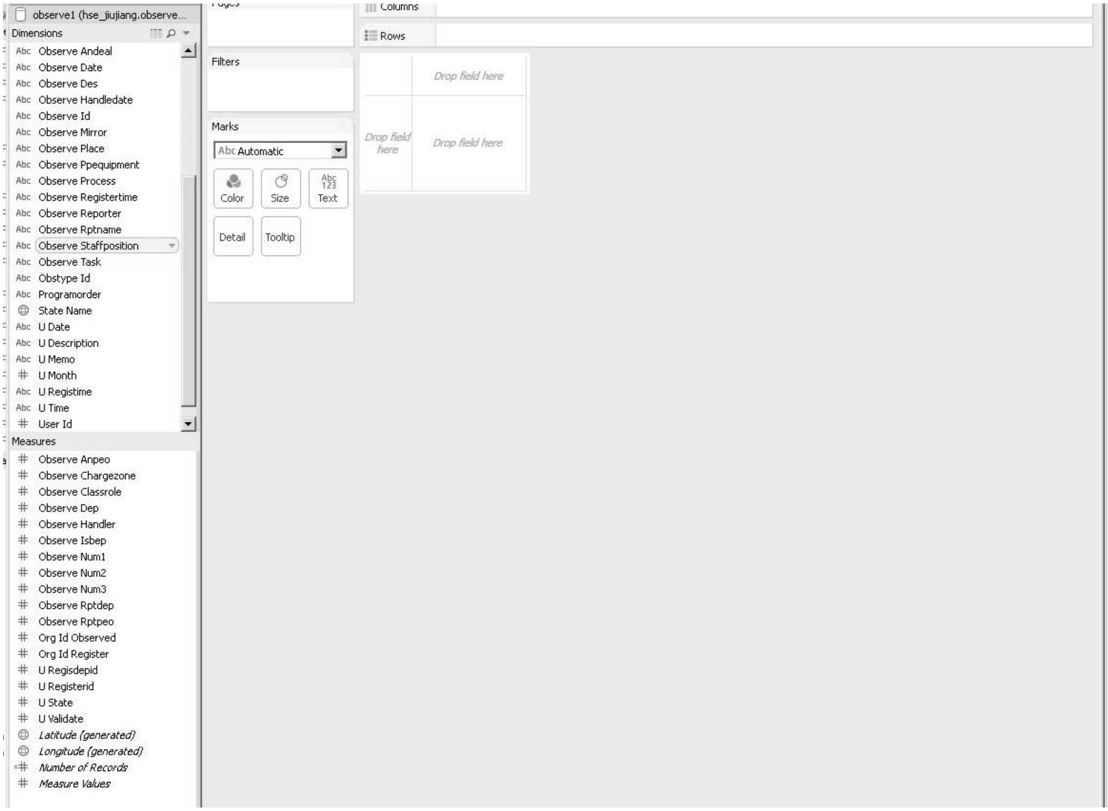


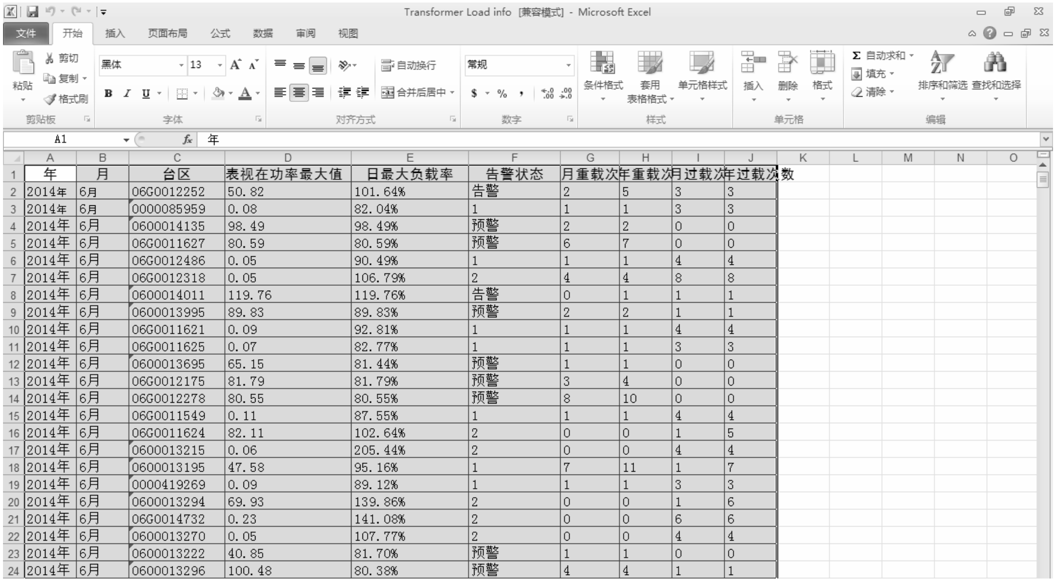
图3-20　连接Hadoop数据示例

3.2.3　复制粘贴输入数据

创建数据源的另外一种方式是将数据复制粘贴到Tableau中，Tableau会根据复制数据自动创建数据源。用户可以直接复制的数据包Microsoft Excel和Word在内的Office应用程序数据、网页中HTML格式的表格、用逗号或制表符分隔的文本文件数据。

下面我们以本地某个Excel文件为例来进行说明。

(1) 打开本地“Transformer Load info”Excel文件，并复制所有数据，如图3-21所示。



	年	月	台区	表视在功率最大值	日最大负载率	告警状态	月重载次	年重载次	月过载次	年过载次数
2	2014年	6月	06G0012252	50.82	101.64%	告警	2	5	3	3
3	2014年	6月	0000085959	0.08	82.04%	1	1	1	3	3
4	2014年	6月	0600014135	98.49	98.49%	预警	2	2	0	0
5	2014年	6月	06G0011627	80.59	80.59%	预警	6	7	0	0
6	2014年	6月	06G0012486	0.05	90.49%	1	1	1	4	4
7	2014年	6月	06G0012318	0.05	106.79%	2	4	4	8	8
8	2014年	6月	0600014011	119.76	119.76%	告警	0	1	1	1
9	2014年	6月	0600013995	89.83	89.83%	预警	2	2	1	1
10	2014年	6月	06G0011621	0.09	92.81%	1	1	1	4	4
11	2014年	6月	06G0011625	0.07	82.77%	1	1	1	3	3
12	2014年	6月	0600013695	65.15	81.44%	预警	1	1	0	0
13	2014年	6月	06G0012175	81.79	81.79%	预警	3	4	0	0
14	2014年	6月	06G0012278	80.55	80.55%	预警	8	10	0	0
15	2014年	6月	06G0011549	0.11	87.55%	1	1	1	4	4
16	2014年	6月	06G0011624	82.11	102.64%	2	0	0	1	5
17	2014年	6月	0600013215	0.06	205.44%	2	0	0	4	4
18	2014年	6月	0600013195	47.58	95.16%	1	7	11	1	7
19	2014年	6月	0000419269	0.09	89.12%	1	1	1	3	3
20	2014年	6月	0600013294	69.93	139.86%	2	0	0	1	6
21	2014年	6月	06G0014732	0.23	141.08%	2	0	0	6	6
22	2014年	6月	0600013270	0.05	107.77%	2	0	0	4	4
23	2014年	6月	0600013222	40.85	81.70%	预警	1	1	0	0
24	2014年	6月	0600013296	100.48	80.38%	预警	4	4	1	1

图3-21　复制数据

(2) 打开 Tableau Desktop进入工作区界面，使用快捷方式Ctrl+V，如图3-22所示。



图3-22　粘贴数据

(3) 从主页面可以看到，通过复制粘贴操作连接到了名称以“Clipboard”开头的新数据源。查看该数据源的属性，可以看到通过复制粘贴操作，在临时文件夹中创建了一个以“Clipboard”开头的文本文件，并连接到了该文本数据源，如图3-23所示。



图3-23 以剪贴板复制方式连接数据

用户最初可能选择了本地数据源,后续需要使用服务器数据源进行替代,并关闭初始的数据源。

3.2.4 筛选数据

直接使用数据源的全量数据，在视图设计时可能会导致工作表响应迟缓。如果仅希望对部分数据进行分析，可以使用数据源筛选器。Tableau可以在新建数据源时选择筛选器，也可以在完成数据连接后，对数据源添加筛选器。

1. 在数据连接时应用筛选器

在新建数据源的界面中，选择“筛选器”下方的“添加”，如图3-24所示。

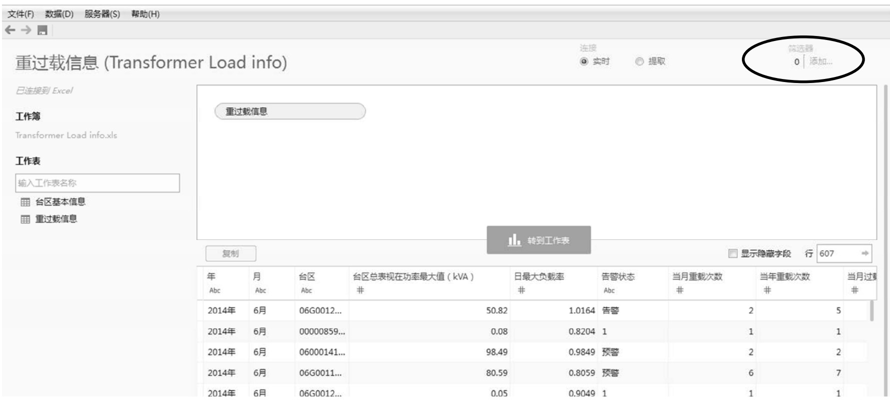


图3-24 选择“添加筛选器”

在“编辑数据源筛选器”对话框中选择“添加”，随即进入“添加筛选器”对话框。例如，选择“日最大负载率”作为筛选字段，如图3-25所示。

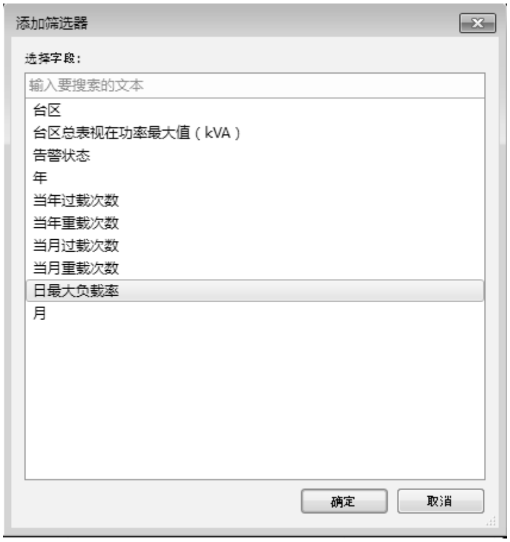


图3-25　添加“日最大负载率”字段

如果只需要分析最大负载率在1.1之上的台区，则需要在“筛选器[日最大负载率]”对话框中将负载率最小值设定为1.1，如图3-26所示。



图3-26　添加数据源筛选器

如果只想分析6月份的最大负载率情况，则在回到“编辑数据源筛选器”对话框后再选择“月”，在“筛选器[月]”对话框中选择“6月”。完成上述操作后，回到“编辑数据源筛选器”对话框，可以看到数据源的筛选字段，如图3-27所示。

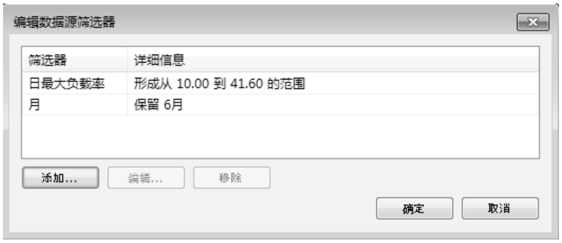


图3-27　完成添加字段筛选

单击“确定”，回到“编辑数据源”界面可以预览筛选后的数据，如图3-28所示。

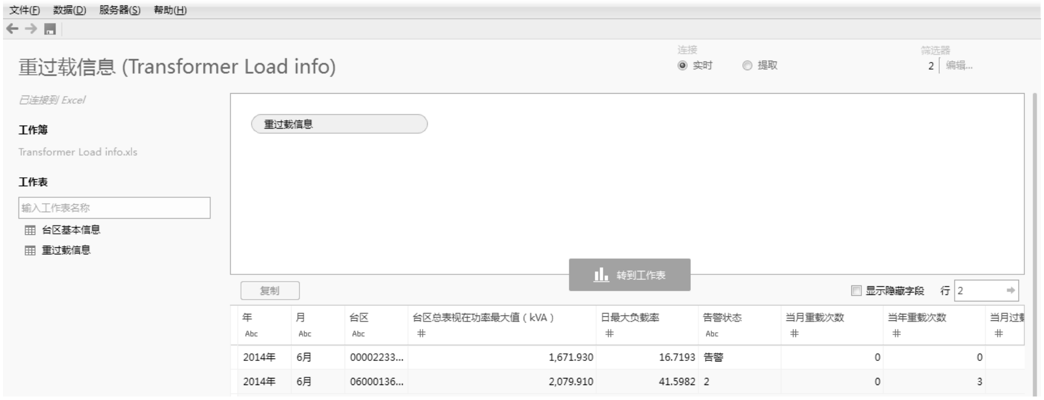


图3-28　筛选后的数据

2. 针对数据源应用筛选器

在完成数据连接后，可以选择“数据”➤“<数据源名称>”➤“编辑数据源筛选器”，后续步骤与在数据连接时应用筛选器的步骤基本一致，这里不再赘述。

3.3　数据整合

上一节的示例中主要说明了数据来自同一张数据表的情况，在实际分析过程中，数据可能来自多张数据表，也可能来自不同的文件或者服务器。Tableau的数据整合功能可实现同一数据源的多表联结、多个数据源的数据融合，以及针对源数据的行列转换。

3.3.1　实现多表联结

在3.2.1节的例子中，已经添加了“Transformer Load info”数据源中的重过载信息，但是要开展进一步数据分析（例如展示台区重过载的地理分布情况），还需要加入台区的基本信息。

在3.2.1节的例子中，选择“数据”➤“<数据源名称>”➤“编辑数据源”，如图3-29所示。

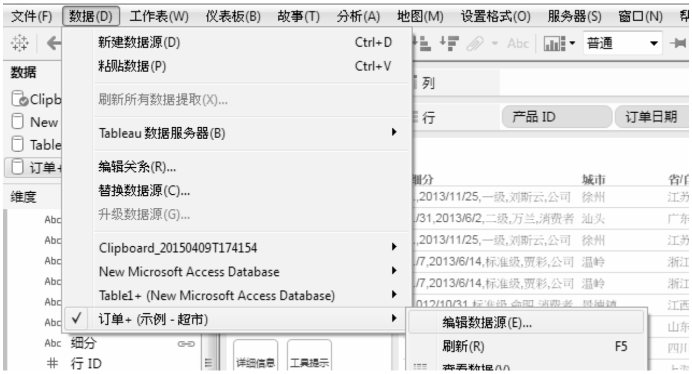


图3-29　打开“编辑数据源”

使用鼠标将“台区基本信息表”拖放到中心区域，Tableau自动将“台区基本信息”与“重过载信息”相联接，如图3-30所示。

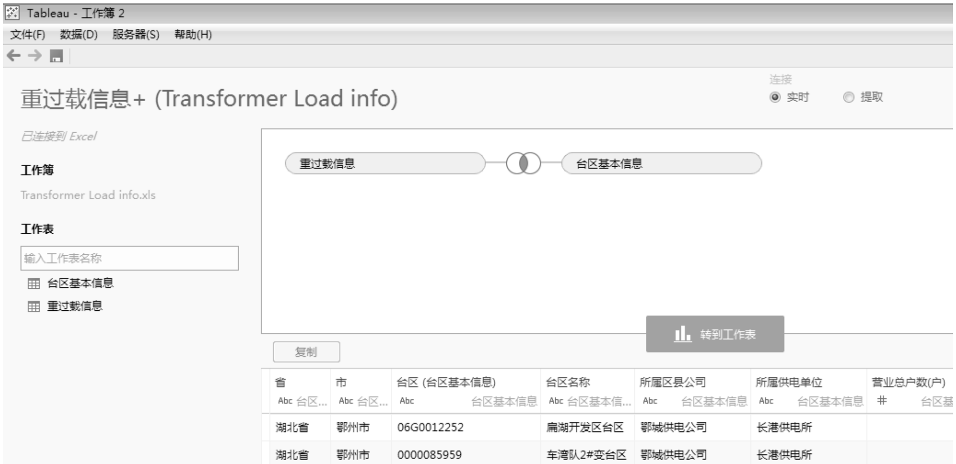


图3-30　添加联接关系

当两表之间无法自动生成表联接时，则显示告警信息，如图3-31所示。



图3-31　无法生成联接示例

如果不希望按照Tableau默认的方式进行表间数据联接，用户可以选择指定表联接方式，操作步骤如下。

单击“联接”图标，可以看到有4种联接类型，默认选择的是“内部联接”，其他选项还包括左侧、右侧、完全外部联接等，如图3-32所示。



图3-32　指定联接方式

说明　联接类型分为内部、左侧、右侧、完全外部4种。其中“内部”只列出与联接条件匹配的数据行；“左侧”表示不仅包含查询结果集合中符合联接条件的行，而且还包括左表的所有数据行；“右侧”表示不仅包含查询结果集合中的符合联接条件的行，而且还包括右表的所有数据行；“完全外部”表示包含查询结果集合中的包含左、右表的所有数据行。

Tableau会自动从多张表中选择同名的字段作为关联字段；如果系统无法自动识别相关对应的字段，也可以手动选择进行关联字段，如图3-33所示。



图3-33　指定关联字段

完成表联接后，选择“转到工作表”，可以在工作区数据窗口中看到“台区基本信息”“重过载信息”两张数据表的信息，如图3-34所示。



图3-34　转到工作区界面

3.3.2　多数据源的数据融合

在3.2.1节的场景里，如果台区的基本信息与重过载信息属于不同的数据源，那么通过多表联接的方式将无法实现台区基本信息和重过载信息的关联。这种情况下，需要通过添加数据源的方式实现数据融合。

在已经建立了“重过载信息”数据连接的基础上，在主界面菜单栏选择“数据”➤“新建数据源”，选择“台区基本信息”文件，添加“台区基本信息”数据源，如图3-35所示。



图3-35　连接到多个数据源

主数据源是在视图中首先使用的数据源。将“重过载信息”数据源的字段拖放到工作区建立视图后，数据源上的标记变为蓝色，表示是主数据源，如图3-36所示。



图3-36　主数据源显示

在主数据源之外，再使用其他数据源后，可以看到数据源用橙色标记，表示此数据源是从数据源，如图3-37所示。



图3-37　从数据源显示

说明　多维数据源（多维数据集）不能用作从数据源，只能用作主数据源。

可以选择“数据”➤“编辑关系”来创建或修改当前数据源关联关系，如图3-38所示。



图3-38　编辑关系

在弹出的“关系”窗口中我们可以通过“主数据源”下拉菜单选择主数据源，如图3-39所示。

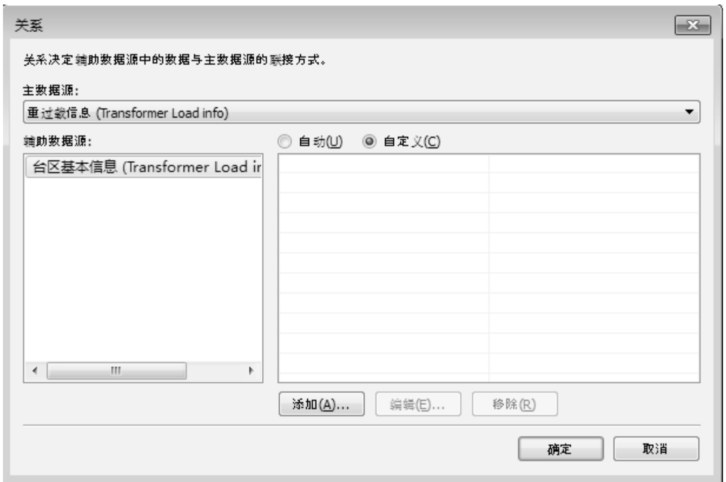


图3-39　“关系”对话框

选择“自定义”，单击“添加”选项来创建新的自定义关系。在弹出的“添加/编辑字段映射”对话框中选择主数据源中的“台区”字段，与从数据源中的“台区”字段进行联接，完成后单击“确定”，如图3-40所示。

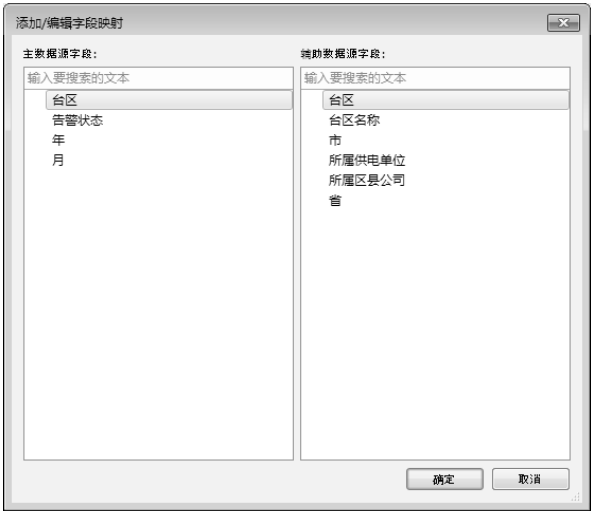


图3-40　创建自定义关系

回到“关系”对话框，可以看到新添加的联接关系，如图3-41所示。

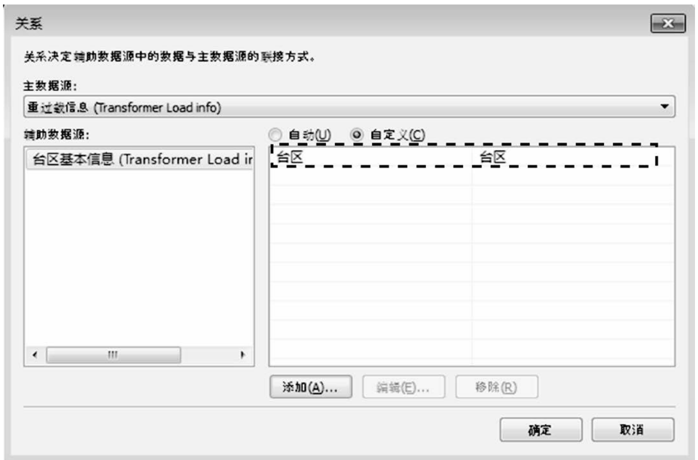


图3-41　自定义添加

在“数据”窗口中，可以看到“台区基本信息”数据源字段和主数据源的联接状态，其中“台区”字段与主数据源“重过载信息”的“台区”字段进行联接，其他字段则不作为联接字段，如图3-42所示。



图3-42 激活的联接字段

联接字段有“激活”和“未激活”两种状态，已激活的联接字段用图标来指示（见图3-42），未激活的联接字段用图标来指示。处于未激活状态时，主、从数据源不使用该联接字段进行联接，单击图标可以进行状态切换。

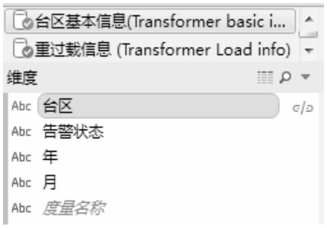


图3-43 未激活的联接字段

3.3.3 行列转换

在使用Tableau进行数据分析时，有时我们需要将源数据中的不同列整合至同一列， Tableau 9.0及以上版本支持对源数据的行列转换。

在数据源窗口，按住Shift或Ctrl，同时选中需要进行转换的列，单击已选择的任一列右侧的按钮，在弹出对话框中选择“数据透视表”，如图3-44所示。



图3-44 执行“行列转换”操作

执行完此操作后可发现，Tableau自动产生新列“数据透视表字段名称”，原列名称转换为该列的不同字段值，同时自动产生新列“数据透视表字段值”，原列的字段值转换至该列，用户可根据实际业务含义对列名进行修改，如图3-45所示。

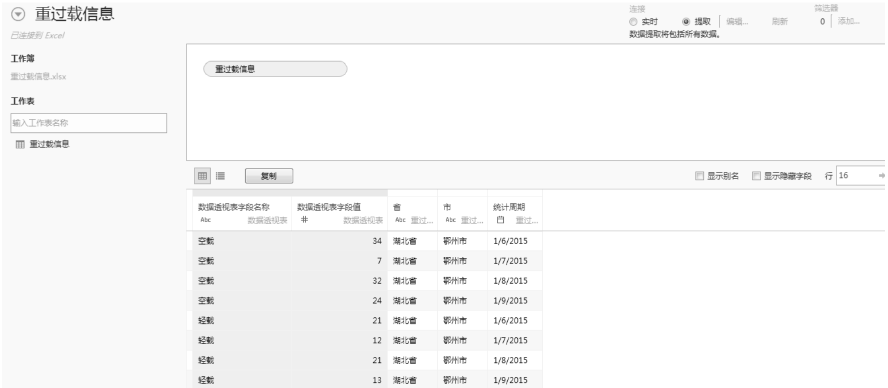


图3-45　“行列转换”操作完成界面

3.4　数据加载

Tableau加载数据有两种基本方式：一种是实时连接，即Tableau从数据源获取查询结果，本身不存储源数据；另一种是数据提取，将数据提取到Tableau的数据引擎中，由Tableau进行管理。本节重点介绍数据提取。

在下列情况下，建议使用数据提取的方式。

- 源数据库的性能不佳：源数据库的性能跟不上分析速度的需要，则可以由Tableau的数据引擎来提供快速交互式分析。
- 需要脱机访问数据：如果需要在差旅途中脱机访问数据，则可以将相关数据提取到本地。
- 减轻源系统的压力：如果源系统是重要的业务系统，那么建议将数据访问转移到本地，以减轻对源系统的压力。

而在下列情况下，则不建议选择数据提取方式。

- 源数据库性能优越：IT基础设施支持快速数据分析，那么不建议进行数据复制。
- 数据的实时性要求高：需要使用实时更新的数据进行分析，则不建议使用数据提取的方式。
- 数据的保密要求高：出于信息安全考虑不希望将数据保存在本地，则不建议进行数据提取。

3.4.1　创建数据提取

Tableau有两种方式创建数据提取：一种是完成数据连接之后，针对数据源进行提取数据操作；

另一种是在新建数据源时选择“提取”方式。

1. 对数据源进行“提取数据”操作

在主界面选择“数据”➤“<数据源名称>”➤“提取数据”，进入提取数据对话框；也可以选择“数据”➤“<数据源名称>”➤“编辑数据源”➤“提取”➤“编辑”，如图3-46所示。



图3-46　进行提取数据操作

在打开的提取数据对话框中可以看到筛选器、聚合、行数3种提取选项，如图3-47所示。



图3-47　提取数据对话框

选择“添加”，弹出添加筛选器对话框，选择用于筛选器的字段，如图3-48所示。

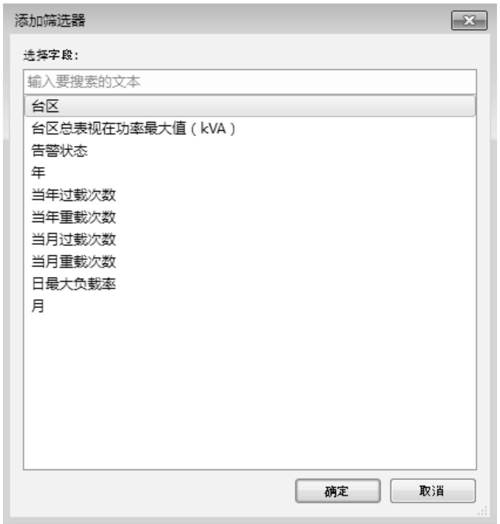


图3-48　选择用作筛选器的字段

可以选择"年"和"月"作为此数据源的数据提取字段，如图3-49所示。



图3-49　选择添加筛选器

在此界面可以指定是否聚合可视维度，也可以选择从数据源提取前若干行。选择"确定"，在弹出的另存为对话框中提取的数据以.tde格式保存，选择"保存"完成创建数据提取，如图3-50所示。

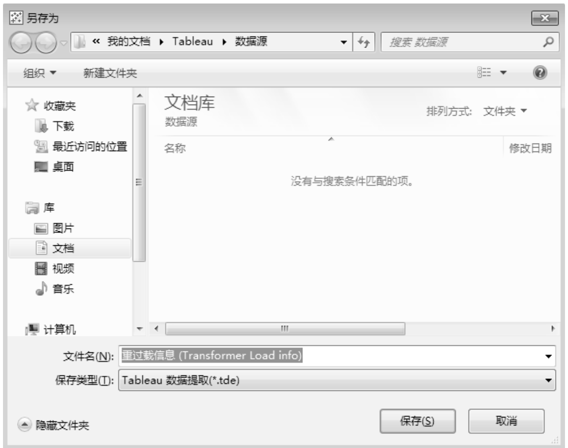


图3-50 提取数据的保存

说明 采用筛选器提取数据时，数据窗口中的所有隐藏字段将会自动从数据提取中排除。单击“隐藏所有未使用的字段”按钮可快速地将这些字段从数据提取中删除。

2. 首次新建数据源时选择“提取”方式

参考3.1节，在新建数据源的过程中，将连接方式从“实时”模式更改为“提取”模式。选择“转到工作表”后，将数据以.tde格式保存，即完成了数据提取的创建，如图3-51所示。

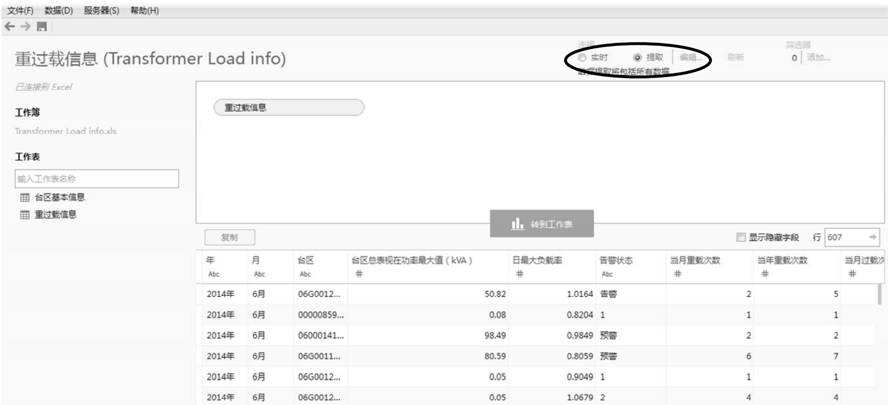


图3-51 选择提取数据

创建数据提取后，当前工作簿开始使用该数据提取中的数据，而不是原始数据源。用户也可以在使用数据提取和使用整个数据源之间进行切换，方法是选择“数据”➤“<数据源名称>”➤“使用数据提取”。

使用数据提取的好处是通过创建一个包含样本数据的数据提取，减少数据量，避免在进行视图设计时长时间等待查询响应，而在视图设计结束后，可以切回到整个数据源。

需要移除数据提取时，可以选择“数据”➤“<数据源名称>”➤“数据提取”➤“移除”。当删除数据提取时，可以选择仅从工作簿删除数据提取，或者删除数据提取文件。后一种情况将会删除在硬盘中的数据提取文件，如图3-52所示。

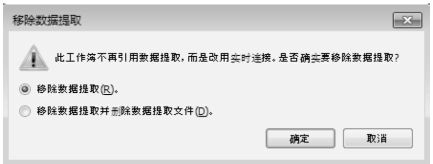


图3-52　移除数据提取

3.4.2　刷新数据提取

当源数据发生改变时，通过刷新数据提取可以保持数据得到更新，方法是“数据”➤“<数据源名称>”➤“刷新”，如图3-53所示。

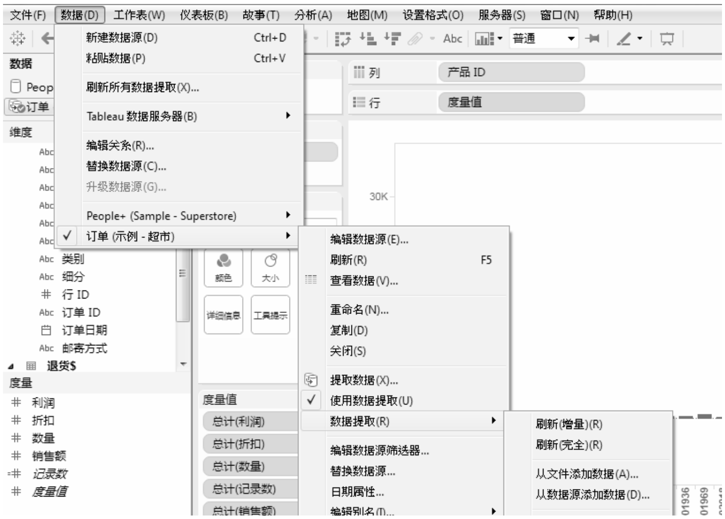


图3-53　刷新数据提取

数据提取的刷新包含两种方式：一种是完全数据提取，即将所有数据替换为基础数据源中的数据；另一种是增量数据提取，仅添加自上次刷新后新增的行。

- **完全数据提取**：默认情况下，将对数据提取进行完全刷新。这意味着每次刷新数据提取时，所有行都会替换为基础数据源中的数据。虽然这种刷新可确保数据提取是数据源的精确副本，但执行这种刷新有时需要大量数据库开销。

- ❑ **增量数据提取：**可以将数据提取设置为仅添加自上次提取数据以来新增的行，而不是每次重新生成整个数据提取。

要改变数据源的提取方式，需要选择“数据”➤“<数据源名称>”➤“提取数据”，如图3-54所示。

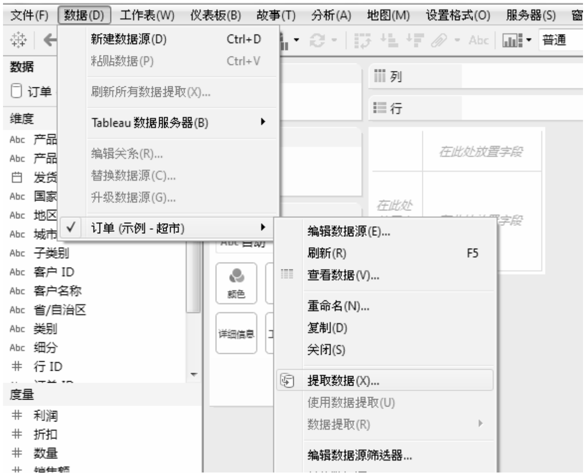


图3-54　打开“数据提取”

在提取数据对话框中，选择“所有行”和“增量刷新”，只有选择提取数据库中的所有行后，才能定义增量刷新，然后在数据库中指定将用于标识新行的字段，如图3-55所示。



图3-55　选择“增量刷新”

用户可以查看刷新数据提取的历史记录，方法是在“数据”菜单中选择数据源，然后选择“数据提取”➤“历史记录”。“数据提取历史记录”对话框将显示每次刷新的时间、类型和所添加的行数，如图3-56所示。



图3-56　查看数据提取历史记录

3.4.3　向数据提取添加行

Tableau可通过两种方式向数据提取文件添加新数据：从文件或从数据源添加。添加新数据行的前提是该文件或数据源中的列必须与数据提取中的列相匹配。

1. 从文件添加数据

当要添加的数据的文件类型与数据提取的文件类型相同时，可以选择从文件数据源向数据提取文件添加新数据。另外一种方式是从Tableau数据提取（.tde）文件添加数据，选择“数据”➤“<数据源名称>”➤“数据提取”➤“从文件添加数据”，如图3-57所示。

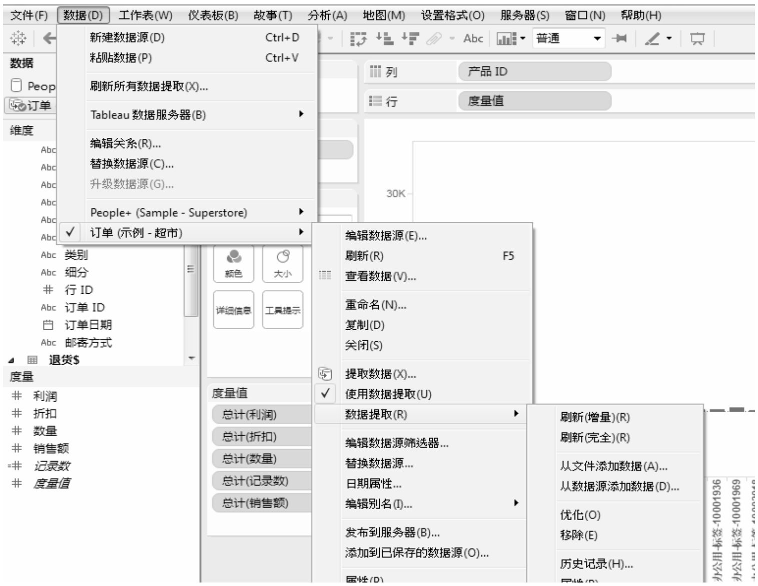


图3-57　从文件添加数据

进入从文件添加数据对话框，选择所要添加的数据文件，单击“打开”，Tableau就会完成从文件添加数据的操作，并提示执行结果，如图3-58所示。



图3-58 成功从文件添加数据

要查看数据添加记录的摘要，可以选择“数据”➤“<数据源名称>”➤“数据提取”➤“历史记录”，如图3-59所示。



图3-59 查看“添加到数据提取”历史记录

2. 从数据源添加数据

另一种添加行的方式，是从工作簿中的其他数据源向所选数据提取文件添加新数据。方法是选择“数据”➤“<数据源名称>”➤“数据提取”➤“从数据源添加数据”。

打开“从数据源追加数据”对话框，选择与目标数据提取文件兼容的数据源，Tableau就会完成从数据源追加数据的操作，并提示执行结果。

3.4.4 优化数据提取

如果要提高数据提取的性能，可以对数据提取进行优化，提高数据提取的查询响应速度。具体操作方法是选择“数据”➤“<数据源名称>”➤“数据提取”➤“优化”（如图3-60所示），采用以下方式进行优化。

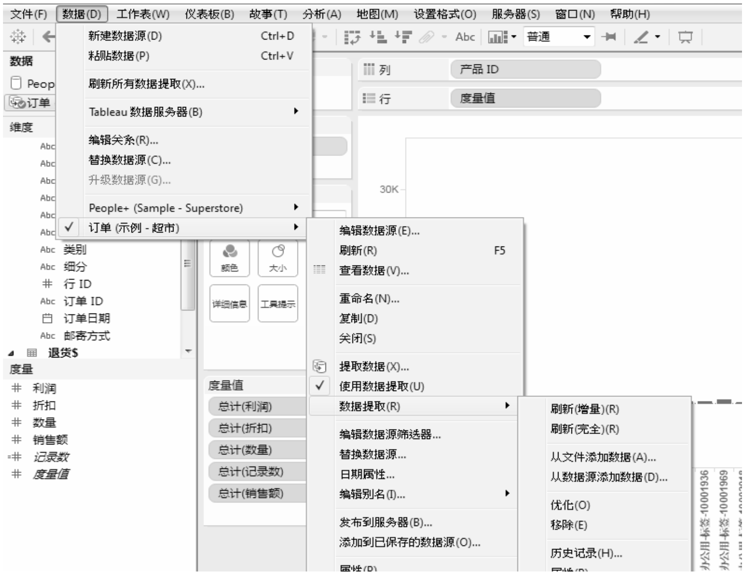


图3-60　优化数据提取

1. 计算字段的预处理

进行数据提取优化后，Tableau提前完成计算字段的预处理，并存储在数据提取文件中。在视图中进行查询时，Tableau可以直接使用计算结果，不必再次计算。

如果改变了计算字段的公式或者删除了计算字段，Tableau将相应地从数据提取中删除计算字段。当再次进行数据提取优化时，Tableau将重新进行计算字段的预处理。

说明　部分函数无法实现预处理，如外部函数（如RAWSQL和R）、表计算函数以及无法提前处理的函数（例如 NOW() 和TODAY()）。

---

2. 加速视图

如果在工作簿内设置了筛选器操作，那么Tableau必须基于源工作表的筛选器，以此计算目标视图的筛选器取值范围。进行数据提取优化后，Tableau将创建一个视图以计算可能的筛选值并缓存这些值，从而提高查询速度。

3.5　数据维护

新建数据源是用户进行数据准备的第一步，在后续工作中，用户需要通过直接查看数据，验证数据连接是否成功；通过添加数据源筛选器，限定分析的数据范围；通过刷新数据源操作，保持分析的数据更新。

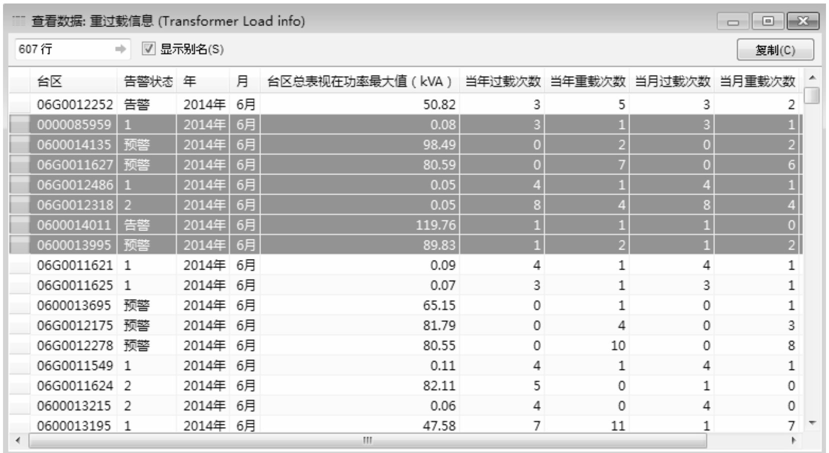
3.5.1 查看数据

查看数据源数据是用户最常见的需求，具体操作方法为选择“数据”➤“<数据源名称>”➤“查看数据”，如图3-61所示。



图3-61 查看数据

在查看数据界面，用户可以选择将数据复制到粘贴板，如图3-62所示。



台区	告警状态	年	月	台区总表视在功率最大值（kVA）	当年过载次数	当年重载次数	当月过载次数	当月重载次数
06G0012252	告警	2014年	6月	50.82	3	5	3	2
0000085959	1	2014年	6月	0.08	3	1	3	1
0600014135	预警	2014年	6月	98.49	0	2	0	2
06G0011627	预警	2014年	6月	80.59	0	7	0	6
06G0012486	1	2014年	6月	0.05	4	1	4	1
06G0012318	2	2014年	6月	0.05	8	4	8	4
0600014011	告警	2014年	6月	119.76	1	1	1	0
0600013995	预警	2014年	6月	89.83	1	2	1	2
06G0011621	1	2014年	6月	0.09	4	1	4	1
06G0011625	1	2014年	6月	0.07	3	1	3	1
0600013695	预警	2014年	6月	65.15	0	1	0	1
06G0012175	预警	2014年	6月	81.79	0	4	0	3
06G0012278	预警	2014年	6月	80.55	0	10	0	8
06G0011549	1	2014年	6月	0.11	4	1	4	1
06G0011624	2	2014年	6月	82.11	5	0	1	0
0600013215	2	2014年	6月	0.06	4	0	4	0
0600013195	1	2014年	6月	47.58	7	11	1	7

图3-62 查看数据界面

3.5.2 刷新数据

当数据源中的数据发生变化后（包括添加新字段或行、更改数据值或字段名称、删除数据或字段），需要重新执行新建数据源操作，才能反映这些修改；另外，也可以执行刷新操作，在不断开连接的情况下即时更新数据，如图3-63所示。

3

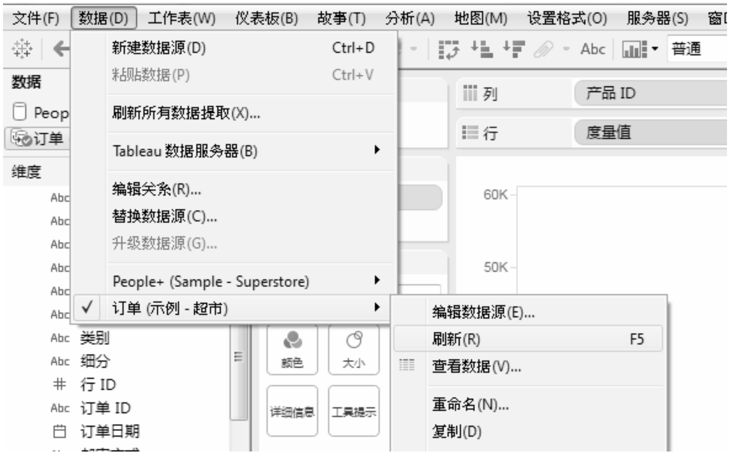


图3-63　刷新数据源

说明　如果工作簿中视图所使用的数据源字段被移除，那么完成刷新数据操作后，将显示一条警告消息，说明该字段将从视图中移除。由于缺少该字段，工作表中使用该字段的视图将无法正确显示。

3.5.3　替换数据

如果希望使用新的数据源(例如Transformer Load Info Update)来替换已有的数据(Transformer Load Info)，而不希望新建工作簿，那么可以进行替换数据源操作。具体方式是选择“数据”➤“替换数据源”，进入替换数据源对话框，如图3-64所示。



图3-64　替换数据源

将原有数据源替换为新数据源，单击“确定”，如图3-65所示。



图3-65 替换操作

完成数据源替换后，当前工作表的主数据源变更为新数据源，如图3-66所示。

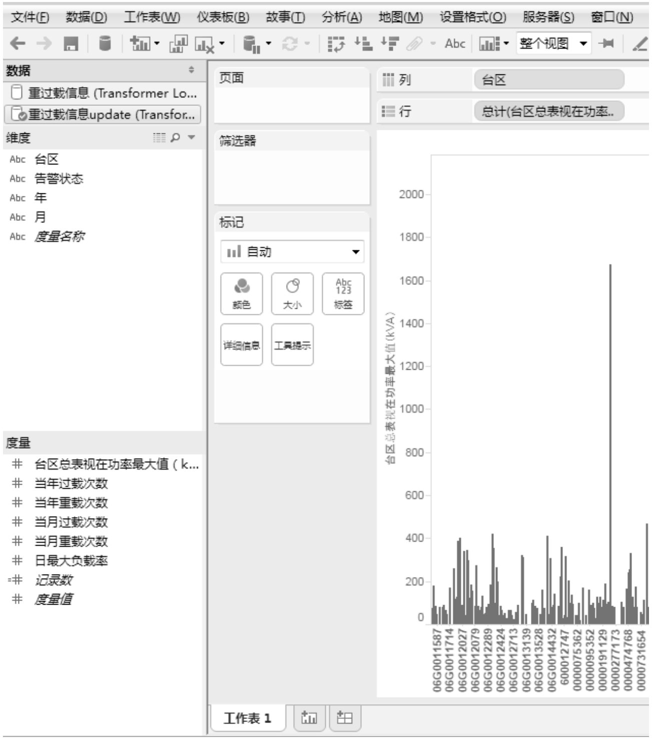


图3-66 数据源替换完成

3.5.4 删除数据

使用了新数据源后，可以关闭原有数据源连接，具体方法是选择“数据”➤“<数据源名称>”➤“关闭”操作来直接关闭数据源，如图3-67所示。



图3-67　关闭数据源

执行关闭数据源操作后，被关闭数据源将从数据源窗口中移除，所有使用了被删除数据源的工作表也将被一同删除。

第 4 章

初级可视化分析

本章将以电力行业数据及常见可视化分析需求为例，介绍11种初级视图的创建用法，分别是：以“2014年各省市售电量明细表”作为数据源的条形图（4.1）、饼图（4.3）、折线图（4.4）、气泡图（4.8）、圆视图（4.9）、标靶图（4.10）；以“公司年龄统计表”作为数据源的直方图（4.2）；以“2014年上半年综合计划指标明细表”作为数据源的基本表（4.5）、压力图（4.6）、树地图（4.7）；以“物资采购情况明细表”作为数据源的甘特图（4.11）。通过本章，读者可以学习创建各类初级视图的操作过程和使用它们进行可视化分析的方法。

4.1 条形图

条形图，又称条状图、柱状图、柱形图，是最常使用的图表类型之一，它通过垂直或水平的条形展示维度字段的分布情况。水平方向的条形图即为一般意义上的条形图，垂直方向的条形图通常称为柱形图。条形图最适宜比较不同类别的大小，需注意纵轴应从0开始，否则很容易产生误导。

本节将介绍如何创建一个用于查看2014年3月份各省市售电量对比的水平/垂直条形图，步骤如下。

(1) 将“统计周期”字段拖到筛选器，如图4-1所示，在弹出的筛选器字段对话框中选择日期类型为年/月，单击“下一步”，再在弹出的对话框中勾选“2014年3月”，这样就把统计周期限制为2014年3月，单击“确定”。



图4-1　添加筛选器，将时间限定为2014年3月

(2) 将维度“省市”拖至行功能区，度量“当期值”拖至列功能区，生成如图4-2所示的视图。

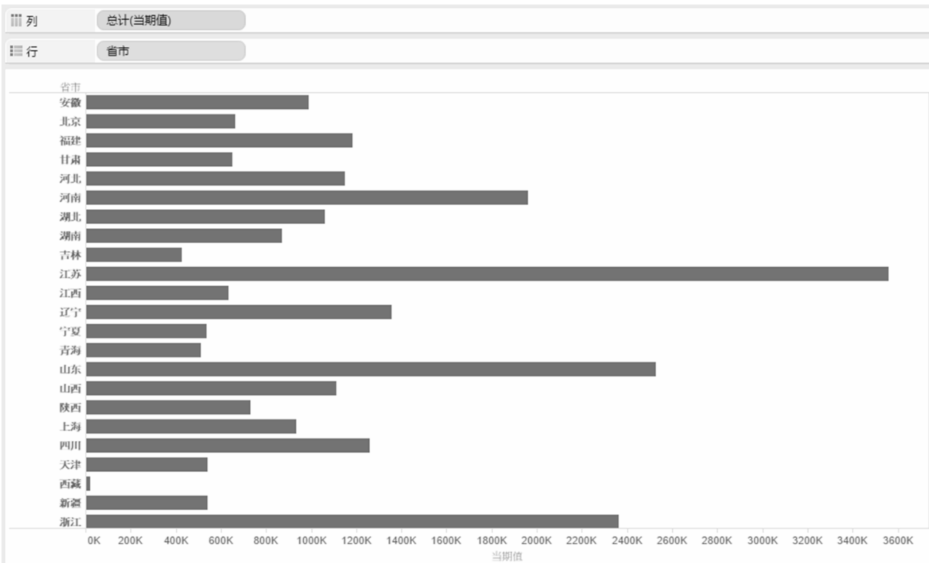


图4-2　2014年3月各省市售电量对比图

(3) 单击工具栏中的“交换”按钮，将水平条形图转置为垂直条形图，单击降序排序按钮，各省售电量将按降序排列，如图4-3所示。

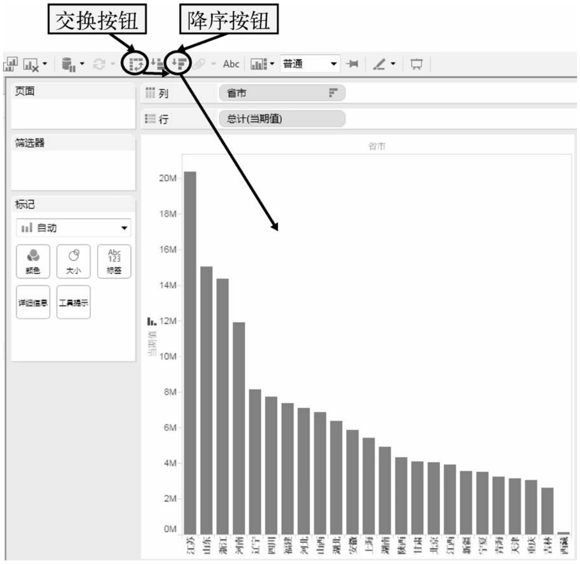


图4-3 对水平条形图进行交换、降序排列

由该视图可以看出，2014年3月，售电量最高的3个省分别为：江苏、山东和浙江，最低的3个省分别为：重庆、吉林和西藏。

(4) 将维度“用电类别”拖至标记卡上的“颜色”，生成堆积条形图，继续查看各省市售电量按用电类别的分布情况，如图4-4所示。

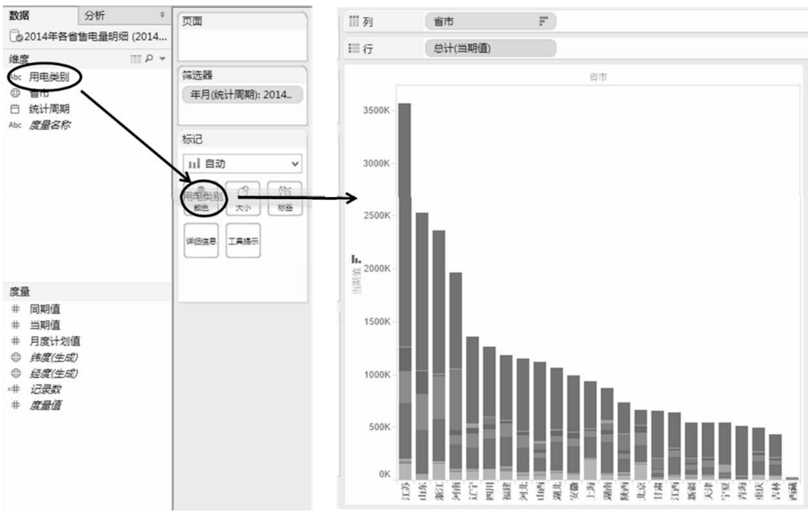


图4-4 2014年3月各省市售电量用电类别分布图

可以发现，当省市维度字段的成员过多时，生成的堆积条形图不够直观，可对堆积条形图中

各用电类别电量进行升降排序。

(5) 单击“用电类别”图例卡的下拉菜单按钮，选择“排序…”，在出现的排序窗口中对排序进行设置。窗口中显示有多种排序方式，包括升序、降序以及升降排序的依据，此外，还可以手动编辑顺序等。这里我们选择按字段“当期值”总计的升序排列，如图4-5所示。

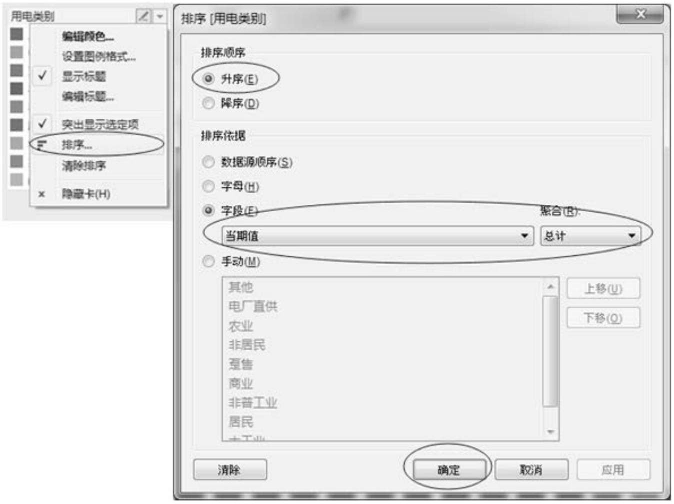


图4-5　为堆积条形图设置排序

设置完成后，颜色的顺序将按照用电类别当期值总计的大小按升序排列（如图4-6所示）；此外，为使视图颜色更好看，还可对各用电类别颜色进行编辑，编辑方式参见2.3.2节。

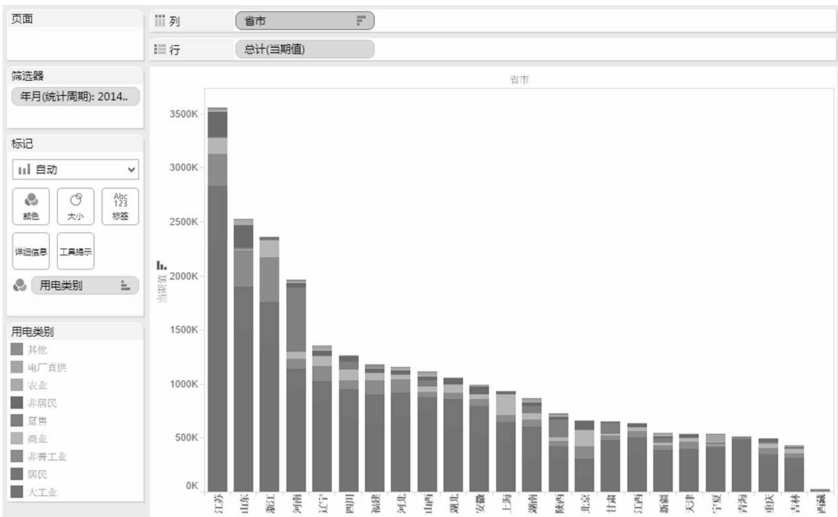


图4-6　排序后的2014年3月各省市售电量用电类别分布图（另见彩插图4-6）

该视图的用电类别已按当期值大小排序，可以清晰地看出，2014年3月，几乎各省市的主要用电类别均为“大工业”“居民”和“非普工业”。

4.2 直方图

直方图与条形图类似，主要区别在于条形图的横轴为单个类别，不用考虑纵轴上的度量值，用条形的长度表示各类别数量的多少；而直方图的横轴为对分析类别的分组（Tableau中称为分级bin），横轴宽度表示各组的组距，纵轴代表每级样本数量的多少。

由此看出，条形图用于展示不同类别的数据时，类别是离散的、较少的，而直方图则是对此类别再进行分组统计。分组的原因可能是因为类别是连续的，或者类别虽然离散但是数量过多，可以视为近似于连续，当然也可以基于某种业务需要。使用直方图分析的样本数据量最好在50个以上。

例如比较分析某公司员工的年龄结构，可考虑将年龄分级为不同的年龄组，再对各年龄组的员工人数进行统计，具体步骤如下。

(1) 加载“公司年龄统计表”作为数据源并创建级。在数据窗口中选择度量“年龄”，单击鼠标右键，在菜单中选择“创建”➤“数据桶...”。在弹出的“编辑级”窗口中，编辑新字段的名称和组距。为帮助确定最佳组距，按“加载”显示值范围，包括最大值、最小值和差异（最大值–最小值）。值范围可以帮助调整设定数据桶大小（Tableau默认为10），数据桶大小也就是直方图中常说的组距，如图4-7所示。



图4-7 创建级字段

因对度量分级创建的该字段为维度字段，故该级字段显示在数据窗口的维度区域中，并在字段名称前附有字段图标，如图4-8所示。

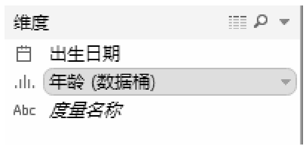


图4-8 创建的级字段显示在维度窗口中

(2) 将度量“记录数”拖至行功能区，将新建的级字段“年龄组”拖至列功能区，生成如图4-9所示的视图。

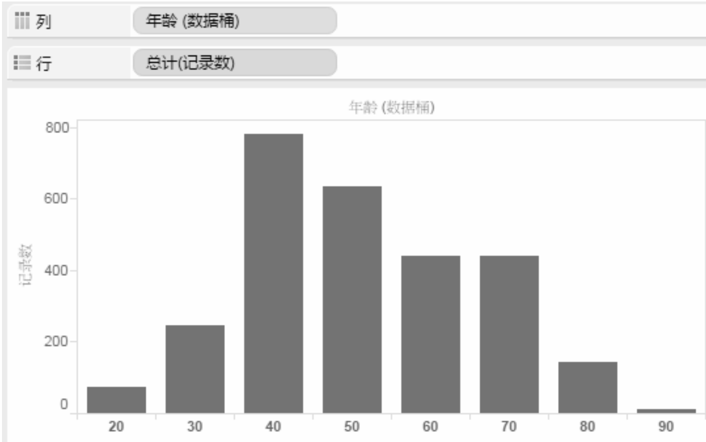


图4-9　年龄分组统计直方图

图4-9中，每个级标签代表的是该级所分配的数字范围的下限（含下限）。例如，标签为30的级的含义是：大于或等于30岁但小于40岁的年龄组。

说明　还可以自动创建直方图，方法是①在数据窗口中选择一个度量；②单击工具栏上的“智能显示”按钮；③选择直方图选项。

(3) 为各级编辑别名。因为自动生成的级仅显示该级的下限，容易产生误导。以修改“20”的标签为例，右键选中“20”级标签，选中“编辑别名”，修改为“20~29”，如图4-10所示。



图4-10　为级标签编辑别名

为每个级标签编辑别名后，最终得到图4-11所示的视图。

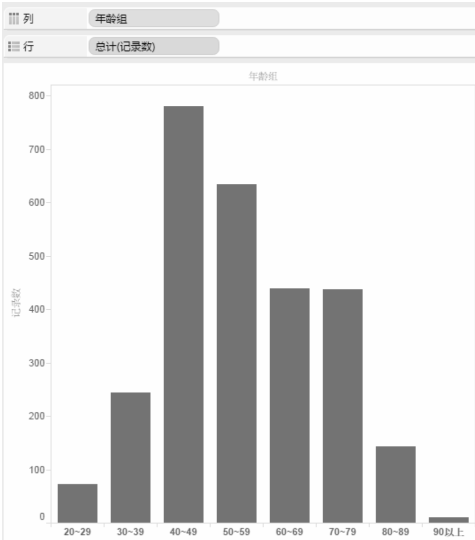


图4-11　编辑后的直方图

由图4-11可以看出，该公司员工的年龄层主要集中在40岁～59岁。

高级应用

运用直方图可以对数据进行分组统计，它同时也是一种非常有效的观察数据分布情况的方法。然而，有时基于业务需要等原因，可以考虑对数据重新分级，通过创建自定义字段[①]的方法，创建大小不同的级。

以上面的问题为例，可将年龄组的级分为：35岁（含）以下、36岁～40岁、41岁～45岁、46岁～50岁、51岁～55岁、56岁（含）以上。

自定义新的计算字段“年龄组（不规则分级）”，如图4-12所示。

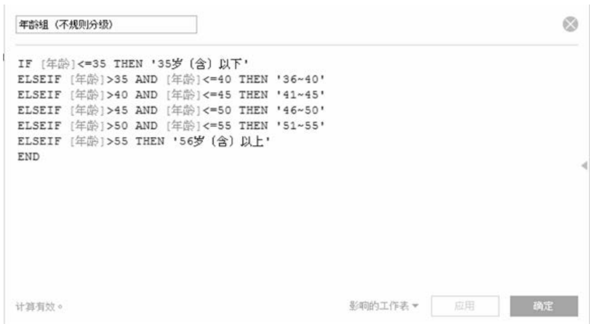


图4-12　创建“年龄组（不规则分级）”

① 关于Tableau自定义字段的详细介绍请参见6.5.2节。

使用该级生成的直方图如图4-13所示。

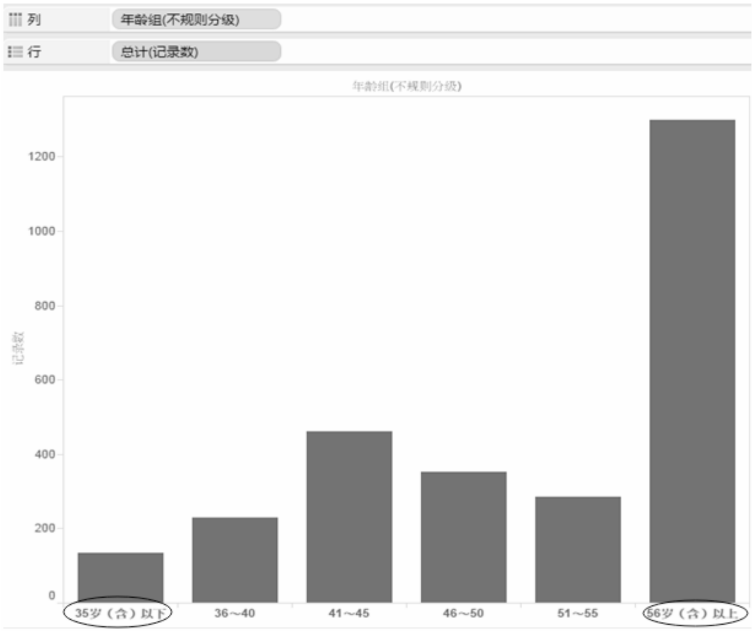


图4-13　使用“年龄组（不规则分级）”生成的直方图

对年龄组进行不规则分级后，我们发现该公司员工的年龄大部分集中在56岁以上。分析出现该现象的原因后发现，该公司为某集团总部，人员多为管理层，因此年龄偏大。此外，还可看出，总部人员老龄化严重，人才年龄结构不健康，需考虑年龄结构调整和后备干部培养等问题。该例说明，有时相等间距的级并不能满足业务的要求，应根据业务特点进行分级，才能更有效地帮助分析人员发现问题。

4.3　饼图

饼图在数据分析中无处不在，然而多数统计学家却对饼图持否定态度。相对于饼图，他们更推荐使用条形图或折线图，因为相对于面积，人们对长度的认识更精确。

注意　使用饼图进行可视化分析时，有许多注意事项，如下所示：

- ❑ 分块越少越好，最好不多于4块，且每块必须足够大；
- ❑ 确保各分块占比的总计是100%；
- ❑ 避免在分块中使用过多标签。

本节将以分析“2014年3月售电量中各用电类别的占比情况”为例，介绍创建饼图的操作步

骤。由于用电类别过多，直接画出这样的饼图不够直观，因此需要利用分组“降低”用电类别的成员数量。具体作图步骤如下。

(1) 从图4-6的结论可知，2014年3月几乎各省市的主要用电类别均为“大工业”“居民”和“非普工业”，因此我们将剩余的用电类别成员归为一组“其他类别”。右键选择维度“用电类别”，选择“创建组...”，按住Ctrl选中要分为一组的成员，单击“分组”即可（详细分组方法可参见6.2节），如图4-14所示。



图4-14 创建“用电类别（组）”

(2) 将字段“统计周期”拖放到筛选器，选择“2014年3月”，将字段“用电类别（组）”拖至“标记”卡的“颜色”，并设置标记类型为“饼图”，“标记”卡中出现“角度”选项，如图4-15所示。



图4-15 用“用电类别（组）”生成饼图后出现“角度”

(3) 将度量“当期值”拖至“角度”后，饼图将根据该度量的数值大小改变饼图扇形角度的大小，从而生成占比图，如图4-16所示。

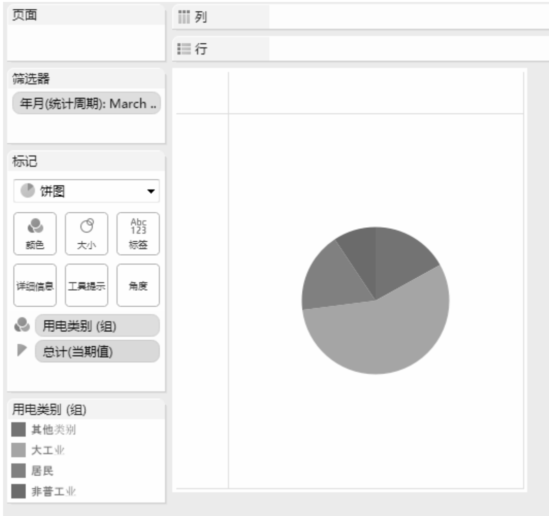


图4-16　2014年3月售电量按用电类别（组）占比饼图

(4) 为饼图添加占比信息。将维度“用电类别（组）”及度量“当期值”拖至“标记”卡中的“标签”，并对标签“当期值”设置“快速表计算”➤“总额百分比”，添加占比信息后的视图如图4-17所示。



图4-17　添加标签后的饼图

4.4　折线图

折线图是一种使用率很高的图形，它是一种以折线的上升或下降来表示统计数量的增减变化趋势的统计图，最适用于时间序列的数据。与条形图相比，折线图不仅可以表示数量的多少，而且可以直观地反映同一事物随时间序列发展变化的趋势。

4.4.1　基本折线图

下面我们以分析“2014年上半年每月售电量的趋势及上年同期情况”为例，介绍创建基本折线图的操作步骤。

(1) 将“当期值”拖至行功能区，“统计周期”拖至列功能区，并通过右键将其日期级别设为“月”，如图4-18所示。

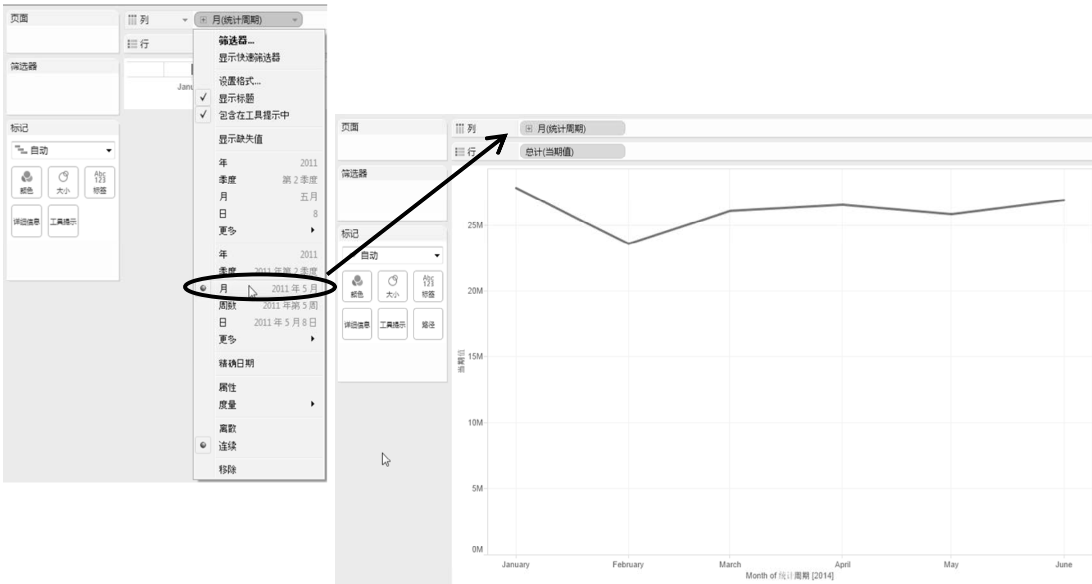


图4-18　2014年上半年每月售电量趋势图

(2) 单击“标记”卡处的颜色，在弹出对话框的标记处选择中间的全部，这时视图中的线段上将出现小圆的标记符号，如图4-19所示。

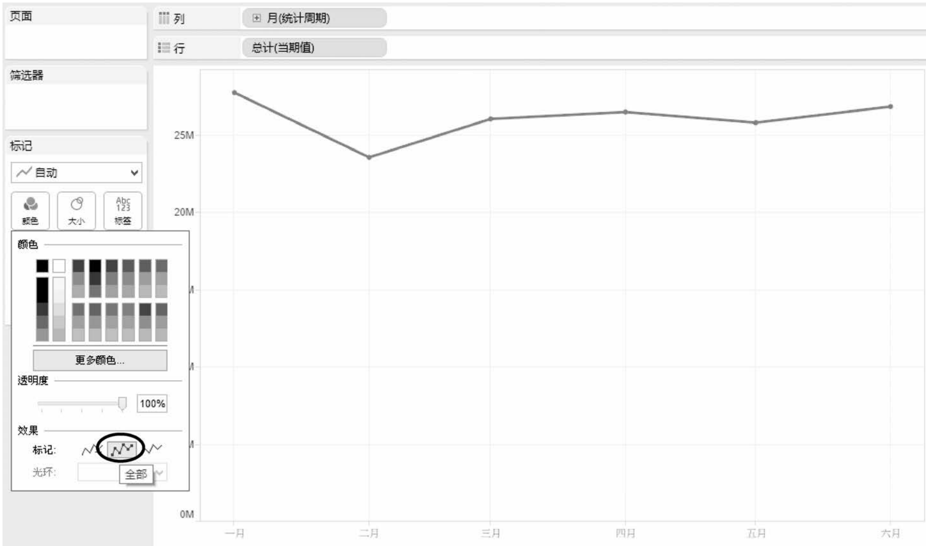


图4-19　为折线图添加标记

有时我们并不满足于标记为一个小圆点，若要标记为一个方形，可以画一个折线图和一个自定义形状的圆图，然后通过双轴来完成，步骤如下。

(1) 两次拖放字段“当期值”到行功能区，这时会出现两个折线图，在“标记”卡处选择其中一个折线图，将标记类型改为形状，如图4-20所示。

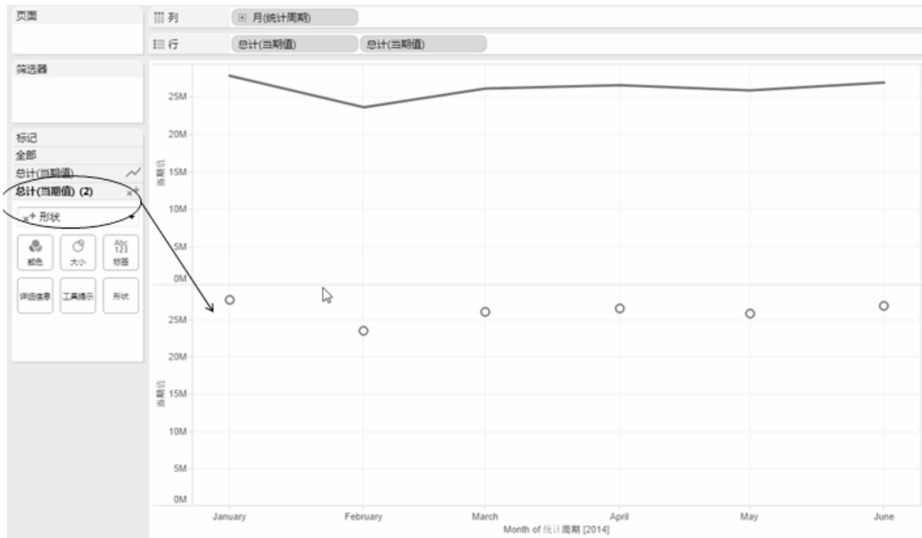


图4-20　创建一个折线图和一个形状图

(2) 单击“标记”卡处的形状，选择方形，可单击大小按钮对方形大小进行调整，如图4-21所示。

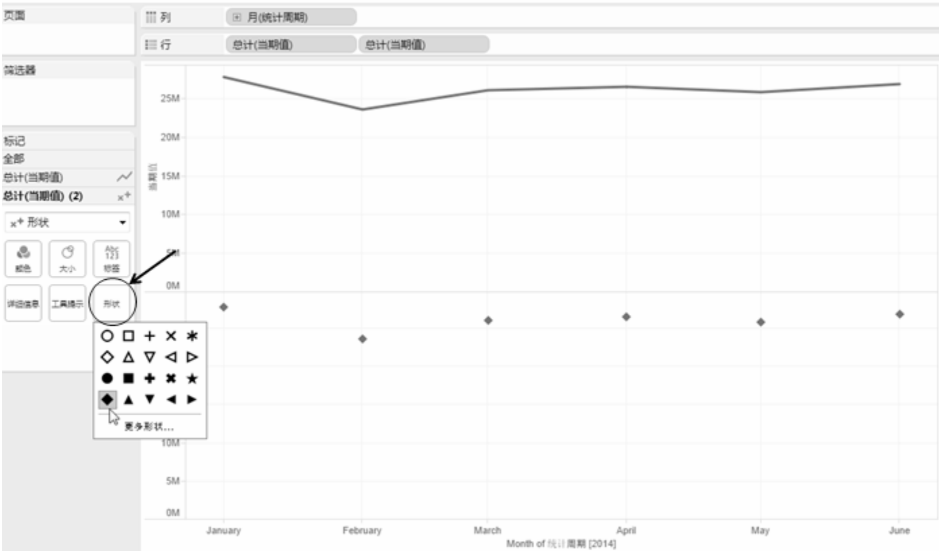


图4-21 定义形状为方形

(3) 右键行功能区右端的“总计（当期值）”，在弹出的对话框中选择双轴。由于两轴的坐标轴均为当期值，因此右键右边的纵轴，选择“同步轴”，完成双轴视图，如图4-22所示。

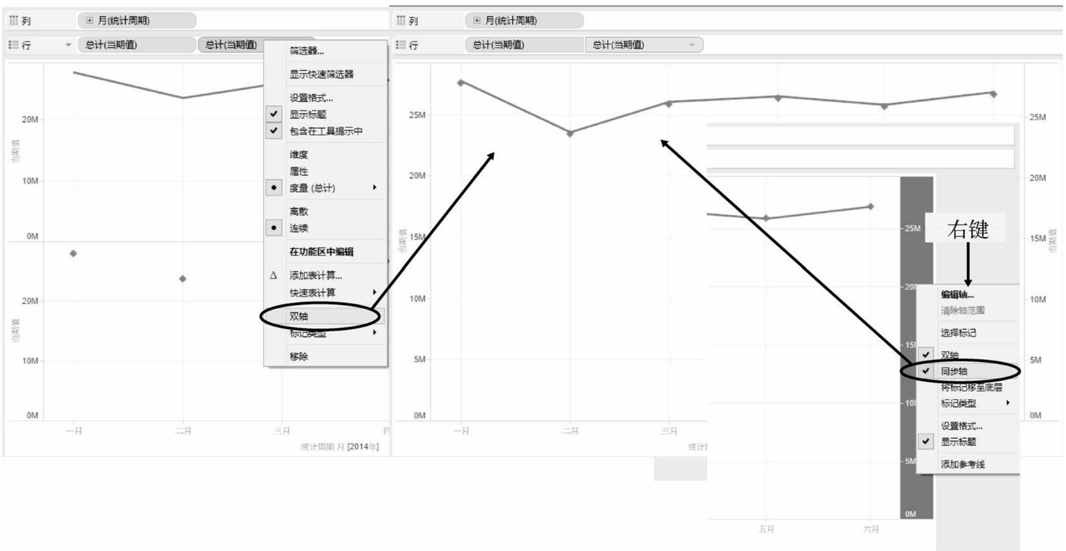


图4-22 双轴创建视图

4.4.2　双组合图

双组合图，又称双轴折线图，是在同一个图表中分别用两个纵轴标记不同数据类型或数据范围的折线图。

下面我们以上例中的数据为例，分析“2014年上半年售电量的月度计划完成情况”，创建该视图的步骤如下。

(1) 同上例，创建2014年上半年每月售电量趋势图。

(2) 创建计算字段“月度计划完成率”，如图4-23所示。



图4-23　创建计算字段“月度计划完成率”

(3) 将度量“月度计划完成率”拖至视图的右侧，如图4-24所示，生成双轴图，如图4-25所示。

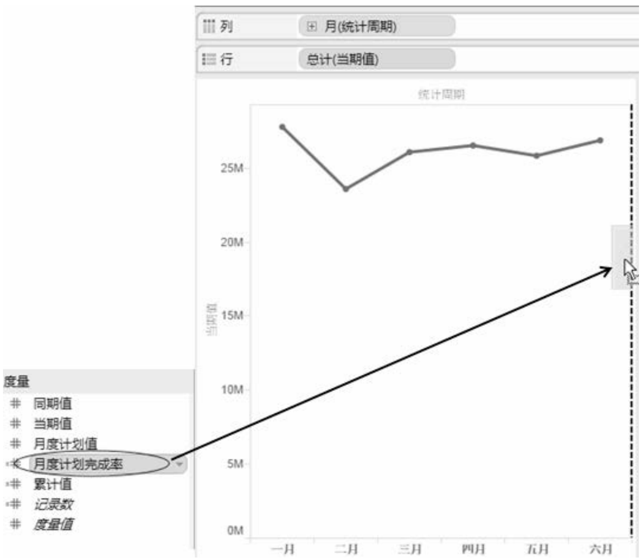


图4-24　将度量拖至视图右侧生成新轴

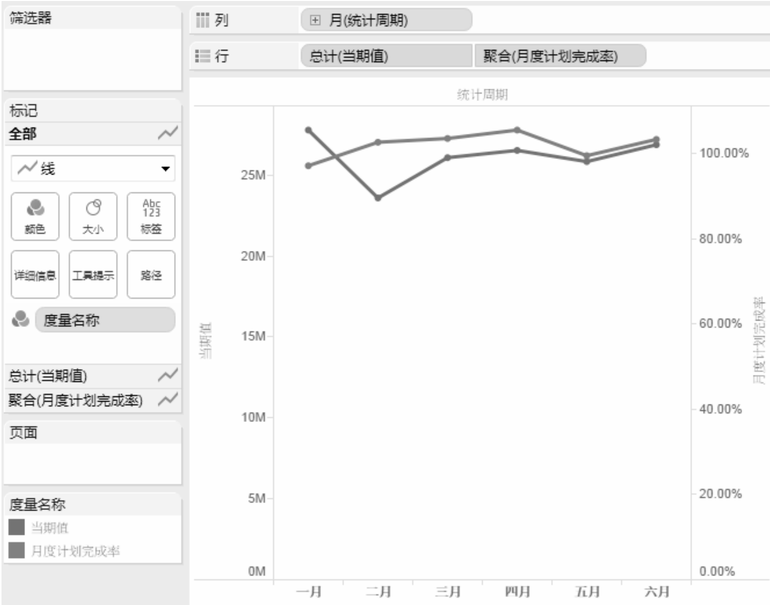


图4-25 双组合图——2014年上半年售电量月度计划完成情况

(4) 因“总计（当期值）”表示数量，可将其标记类型修改为“条形图”，从而生成双组合图，如图4-26所示。

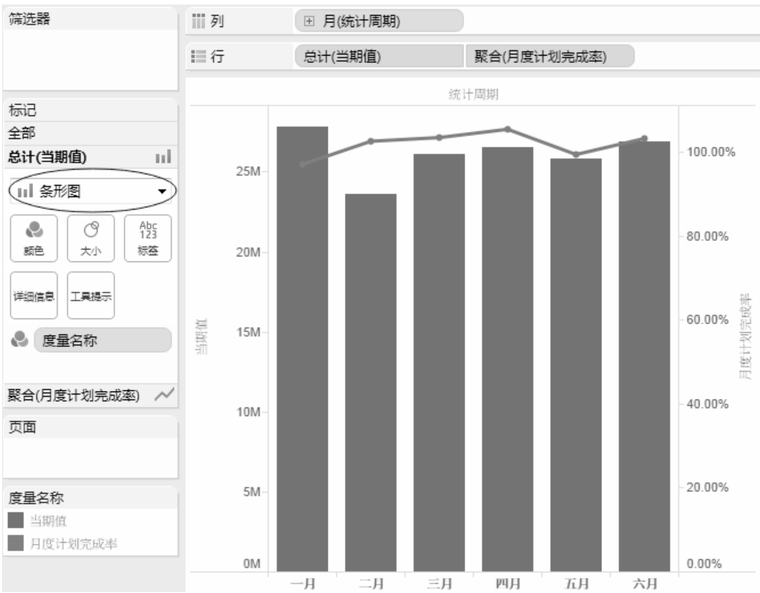


图4-26 双组合图——2014年上半年售电量月度计划完成情况

(5) 为考核月度计划执行率是否达到不低于100%的标准，可以在该视图添加参考线，设置方法如图4-27所示。



图4-27　为轴添加参考线

为“月度计划完成率”添加“标签”，最终生成的视图如图4-28所示。

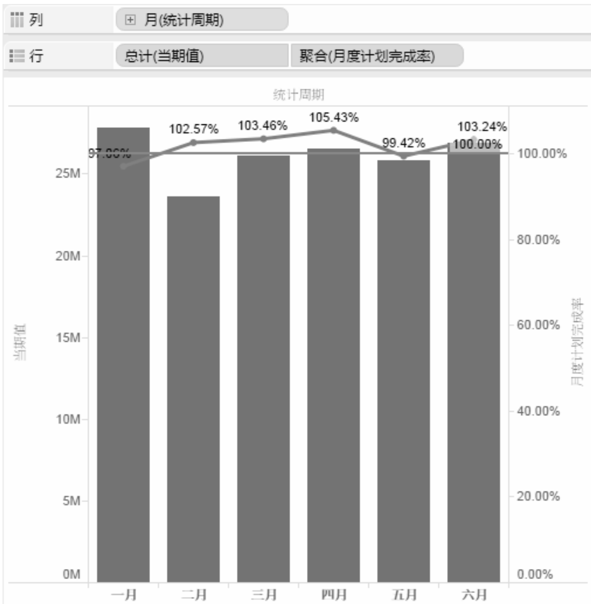


图4-28　添加标签后的双组合图

由图4-28可以看出，2014年上半年，月售电量有两个月没有达到月度计划完成标准，这两个月分别是一月和五月。

4.5　基本表

基本表，又称作文本表、交叉表，即一般意义上的表格，它是一种最为直观的数据表现方式，在数据分析中具有不可忽视的作用。表格可以代替冗长的文字叙述，便于计算、分析和对比。但表格的缺点是不够形象、直观，当表格中数据量较大时，分析人员很难快速定位到所需信息。

下面以分析“2014年上半年全国各省市累计售电量与利润总额情况”为例，介绍创建基本表的方法。

(1) 连接“2014年上半年综合计划指标明细表”数据源，将维度“省市”拖至行功能区，将“指标分组”拖至列功能区，并将度量“当期值”拖至“标记”卡上的“文本”，当期值将自动汇总为2014年上半年累计值，并生成图4-29所示的基本表。



	指标名称	
省市	利润总额	售电量
青海	12,718	3,604,692
陕西	18,306	4,568,282
重庆	17,418	3,243,182
辽宁	91,198	9,145,405
西藏	-42,866	140,441
福建	147,409	8,270,021
甘肃	24,306	4,689,443
湖南	35,431	5,388,956
湖北	61,297	7,155,674
浙江	295,945	15,510,703
河南	24,567	14,751,510
河北	95,316	7,198,229
江西	51,123	4,041,114
江苏	339,211	20,610,117
新疆	46,079	3,512,353
山西	92,754	7,818,308
山东	356,050	16,341,718
安徽	83,213	6,054,995
宁夏	28,682	4,134,415
天津	88,582	3,336,482
四川	74,665	8,053,782
吉林	10,788	3,153,478
北京	61,795	4,880,261
上海	170,775	5,258,451

图4-29　基本表——2014年上半年全国各省市累计售电量与利润总额情况表

(2) 为基本表添加列总计。选择菜单栏的“分析”➤“合计”➤“显示列总计”，如图4-30所示。

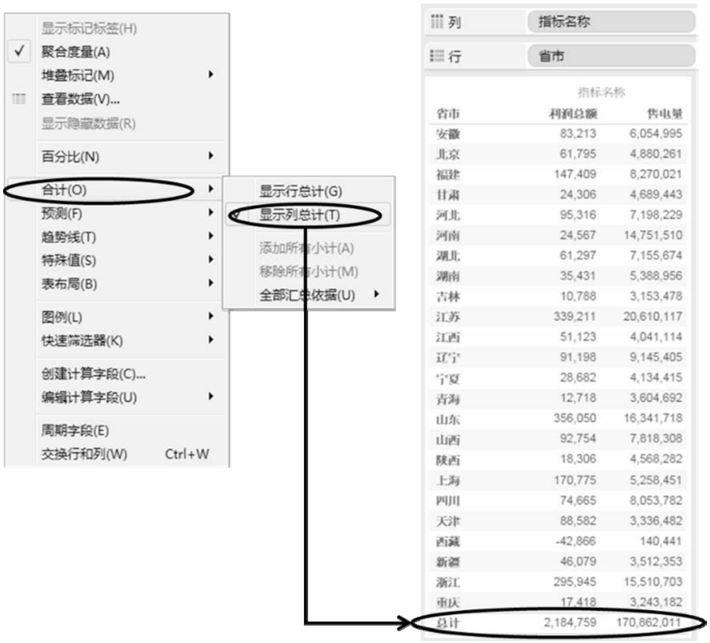


图4-30　为基本表添加列总计

该表格列出了2014年上半年全国各省市累计售电量和利润总额的值，是最直接的一种数据展示方式。

4.6　压力图

由上一节的示例可以看出，当数据量较大时，分析人员很难通过基本表的方式获取重要信息。这时可选择使用压力图（包括突显表）或树形图。

压力图，又称热图、热力图，是一种对表格中数字的可视化表示，通过对较大的数字编码为较深的颜色或较大的尺寸，对较小的数字编码为较浅的颜色或较小的尺寸，可以帮助用户快速地在众多数据中识别异常点或重要数据。当仍需要利用表格展示数据，同时又需要突出重点信息时，可选择使用突显表（参见4.6.2节）。

根据上例的结果，可采用压力图继续分析累计售电量与利润总额的关系，下面以查找其中的异常点为例，介绍压力图的创建方法。

4.6.1　压力图

数据准备：连接“2014年上半年综合计划指标明细表”数据源，创建计算字段“售电量当期值”和“利润总额当期值”（请参考6.5节），计算公式如图 4-31所示。



图4-31　计算字段

(1) 将“省市”拖至行，将“售电量当期值”拖至标记卡的“大小”上，售电量当期值将自动汇总为2014年上半年累计值，并得到图4-32所示的压力图。



图4-32　压力图——2014年上半年各省市累计售电量情况

可以看出，标记的大小代表了售电量的大小，标记越大值越大，标记越小值越小。在图中可以快速地发现重要数据，如江苏、山东和浙江在所有省市售电量中居于前三位。

(2) 可将“利润总额当期值”拖至“标记”卡的“颜色”上，生成图4-33所示的压力图，以快速获取两指标的异常点。



图4-33　压力图——2014上半年各地市累计售电量与利润总额情况（另见彩插图4-33）

由图4-33可以看出，利润总额的大小由颜色表示，绿颜色越深代表利润值越大，相关企业经营成果越好；红色越深代表利润值越小，相应地，企业的亏损情况越严重。图4-33展示的数据即为上例基本表中的数据，但与基本表相比，能够快速地展现两个关联指标的关系以及数据的异常情况。如河南省2014年上半年的售电量位居各省前列，但利润总额却排名靠后；再比如上海随售电量较其他省市比较小，但利润总额情况较好。这一图形结果可以方便分析人员快速定位数据异常点，并可结合对明细的钻取以及实际业务，理解发生异常的原因。

4.6.2　突显表

突显表也是对基本表的一种变形，与压力图类似，目的也是帮助分析人员在大量数据中迅速发现异常情况，但因其显示出具体数值后将与基本表一样，当数据量较大时对异常及重要数据难以辨识，故建议不要用突显表表示相关联指标的情况，而是仅突出显示一个指标（度量）的异常或重要信息。

由上例的结果可知，部分省市虽然售电量较大，但利润情况却不够理想，下面以分析“2014年上半年各地市累计利润总额的情况”为例，介绍创建突显表的方法。

(1) 将“省市”拖至行功能区，将“利润总额当期值”分别拖至“标记”卡的“文本”及“颜色”上，将标记类型改为“方形”，得到图4-34所示的突显表。



图4-34　一个指标的突显表——2014年上半年各省市累计利润总额情况表（另见彩插图4-34）

可以看出，突显表通过各表格颜色的深浅，能够帮助分析人员非常直观、迅速地从大量数据中定位到关键数据，这一点和压力图使用标记大小帮助定位在本质上是相同的；而且突出表还显示了各项的值，又兼具基本表的优点。例如从该例中，分析人员可以快速发现江苏、山东、浙江等地上半年的利润总额位居前列且值分别是多少；而像西藏，可知其利润总额为负值，企业经营亏损。

(2) 使用突显表分析压力图中的例子：查看各省市累计售电量和利润总额中的异常点。将“售电量当期值”拖至“指标卡”中的“颜色”，生成图4-35所示的视图。

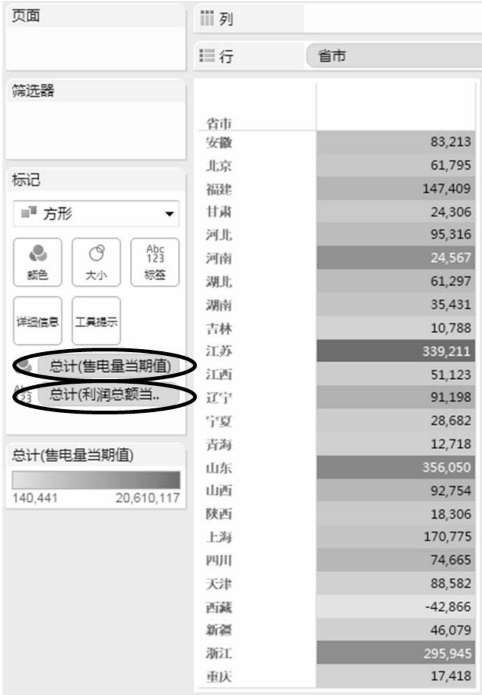


图4-35　两个指标的突显表——2014年上半年各省市累计
售电量与利润总额情况表（另见彩插图4-35）

图4-35中，表格中的数值表示利润总额的大小，单元格的颜色表示售电量的大小。由于利润总额由数值直接表达，传递信息不够直观，因此无法像压力图那样帮助用户快速看出两个相关联指标的异常情况。

通过上例可以发现，突显表在表达关于一个度量“突出值”的情况下是非常有效的。

高级应用

由上例可知，有部分省市的利润总额为负值，假设只想将利润总额为负的数据突出显示，可以进行下列操作。

(1) 将“省市”拖至行功能区，将“利润总额当期值”分别拖至“标记”卡的“文本”及“颜色”上，生成图4-36所示的视图。

(2) 单击“颜色图例”右上角的下拉按钮，选择“编辑颜色”。在弹出的对话框中，选择自定义发散的颜色，并将两端设置为红色和黑色，渐变颜色设定为2，这时变只有红和黑两种颜色。按照分析元素需要，让负数显示为红色，正数显示为黑色，即划分两种颜色的依据是正负，于是单击“高级”，设定中心为0，如图4-37所示。



图4-36　利润总额按数值大小用不同颜色显示（另见彩插图4-36）

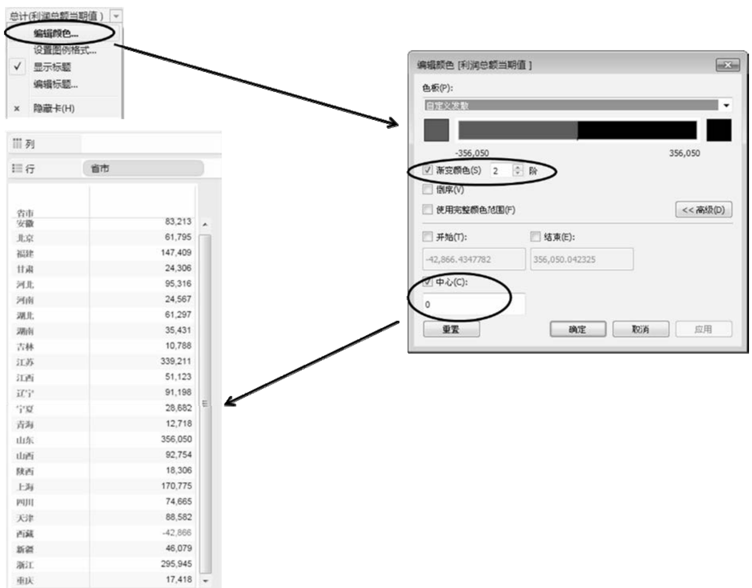


图4-37　突显表——突出显示利润总额的负值（另见彩插图4-37）

由图4-37可以看出，负值已用红色突出显示。

压力图和突显表都可以帮助分析人员快速发现异常数据，并对异常数据进行下钻，从而查看和分析引起异常的原因。

4.7　树地图

树地图，也称树形图，使用一组嵌套矩形来显示数据，同压力图一样，也是一种突出显示异常数据点或重要数据的方法。

同样以分析“2014年上半年全国各省市累计利润总额的关系”为例，创建树地图的方法如下：选择标记类型为方形，将“省市”拖放至“标签”；将“售电量当期值”拖放至“大小”，这时图形的大小代表售电量当期值累计；将“利润总额当期值”拖放至“颜色”，颜色深浅代表大小。最后得到图4-38所示的树形图。

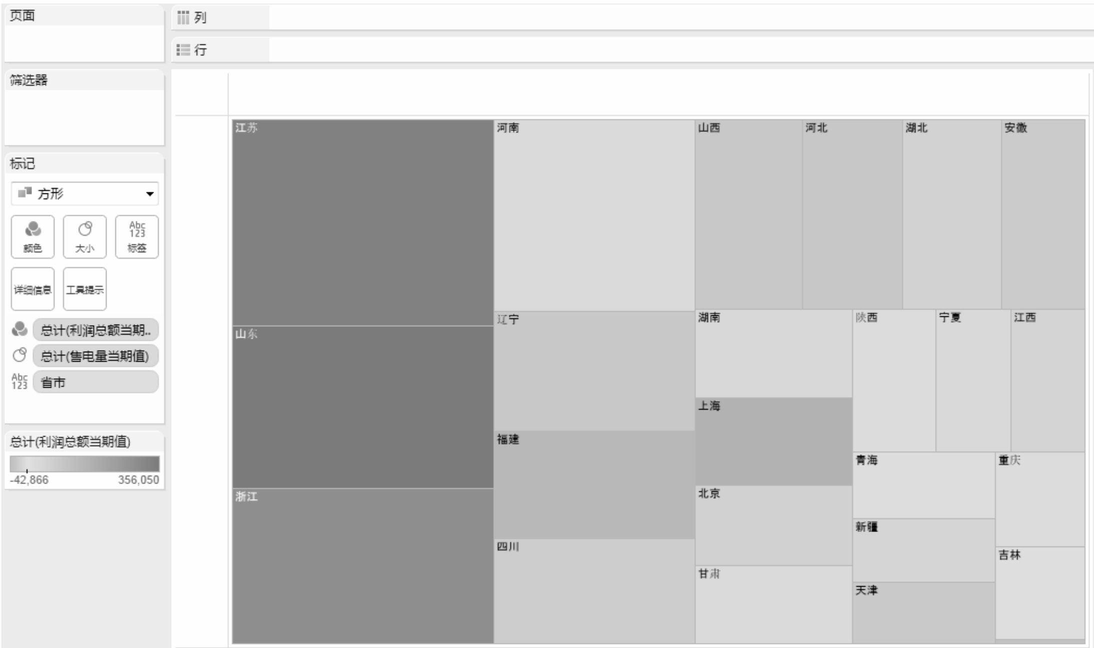


图4-38　树地图——2014年上半年全国各省市累计售电量与利润总额关系（另见彩插图4-38）

图4-38中，矩形的大小代表售电量的大小，颜色的深浅代表利润总额的大小。可以看出，江苏、山东及浙江省上半年累计售电量均排名全国前列；河南、甘肃等地售电量较大但利润情况不佳；而上海的售电量虽小，但利润总额情况较好。

4.8　气泡图

气泡图，即Tableau“智能显示”卡上的“填充气泡图”。每个气泡表示维度字段的一个取值，各个气泡的大小及颜色代表了一个或两个度量的值。Tableau气泡图的特点是具有视觉吸引力，能够以非常直观的方式展示数据。

下面以分析“2014年6月各省市售电量的大小”为例，介绍创建填充气泡图的操作步骤及分析方法。

(1) 加载“2014年各省市售电量明细表”作为数据源并限制“统计周期”为“2014年6月”。将“省市”分别拖至“标记”卡的“颜色”和“标签”，将“当期值”拖至“标记”卡的“大小”，并更改标记类型为“圆”，生成图4-39所示的视图。

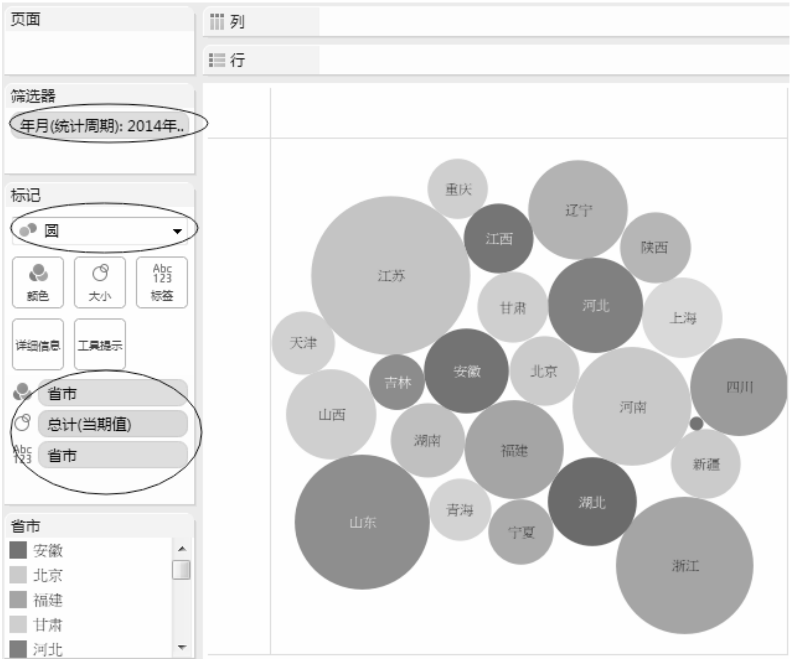


图4-39　填充气泡图——2014年6月各省市售电量情况（另见彩插图4-39）

Tableau会自动用不同的颜色标示出每个省份，并用气泡的大小标示出各省份当月售电量的大小。可以看出，2014年6月江苏省和浙江省的售电量最多。

(2) 将填充气泡图的“标记”由“圆”改为“文本”时，视图将由填充气泡图变为文字云。例如，将上例中“标记”的“圆”改为“文本”后，得到图4-40所示的视图。

由上例可以看出，文字云和填充气泡图的本质相同，但用“文本”的大小替换 “圆”的大小之后，直观性较差。

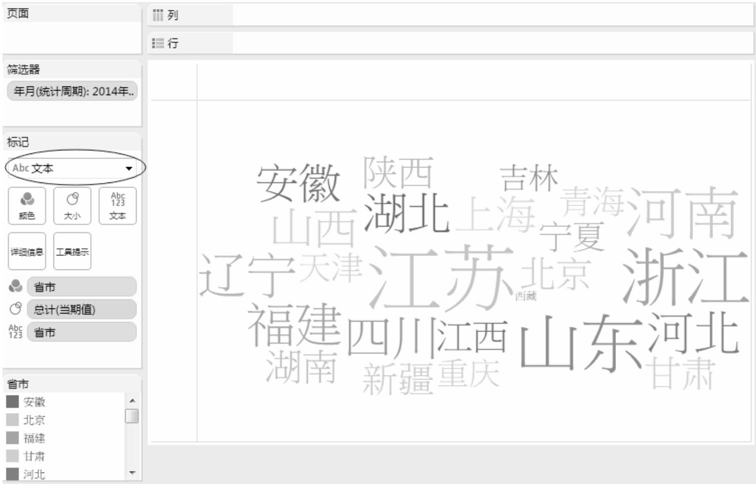


图4-40　文字云——2014年6月各省市售电量情况

4.9　圆视图

圆视图可看作是气泡图的一种变形，通过给气泡图添加一个相关的维度，按不同的类别分析气泡，并依据度量的大小，将所有气泡有序地排列起来，表现较气泡图更为清晰。

下面以分析“2014年上半年售电量按用电类别的各省市分布情况” 为例，介绍圆视图的创建方法。

(1) 将“用电类别”拖至列功能区，将“当期值”拖至行功能区，并修改标记的类型为“形状”，得到2014年上半年累计售电量按用电类别的圆视图，如图4-41所示。

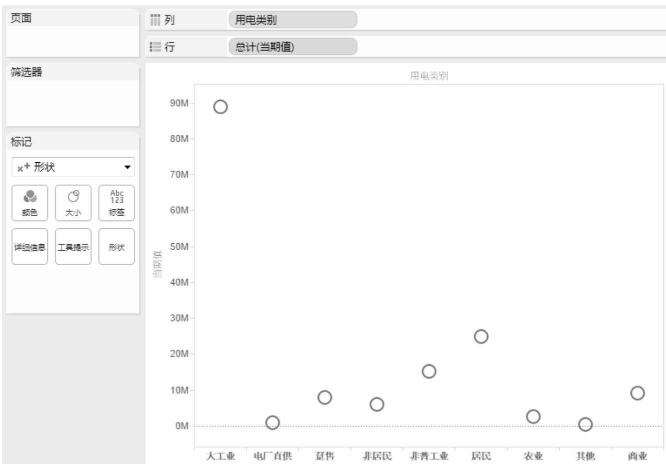


图4-41　圆视图——2014年上半年售电量按用电类别圆视图

(2) 分别将“省市”及“用电类别”拖至“标记”卡上的“大小”和“颜色”，生成图4-42所示的圆视图。



图4-42 圆视图——2014年上半年售电量按用电类别的各省市分布情况

圆视图可以帮助分析人员快速发现每一类别中的异常点或突出数据，例如在图4-42中，很容易就能定位到“用电类别”为“大工业”的江苏省的数据非常突出；再比如“趸售”中河南省的数据较其他各省突出很多。

4.10 标靶图

标靶图是指通过在基本条形图上添加参考线和参考区间，可以帮助分析人员更加直观地了解两个度量之间的关系，常用于比较计划值和实际值。

下面以分析“2014年上半年各省市利润总额的计划完成情况”为例，介绍创建标靶图的操作步骤及分析方法。

(1) 将“省市”拖至行功能区，将“当期值”拖至列功能区，并将“月度计划值”拖放到“标记”卡上，创建图4-43所示的标靶图。

(2) 添加参考线和参考区间。右键单击视图区横轴的任意位置，在弹出菜单上选择“添加参考线”，在弹出的编辑窗口中选择类型为“线”，并对参考线的范围、值及格式进行设置；再对参考区间进行设置，选择类型“分布”，并对范围、区间的取值和格式进行设置，如图4-44所示。

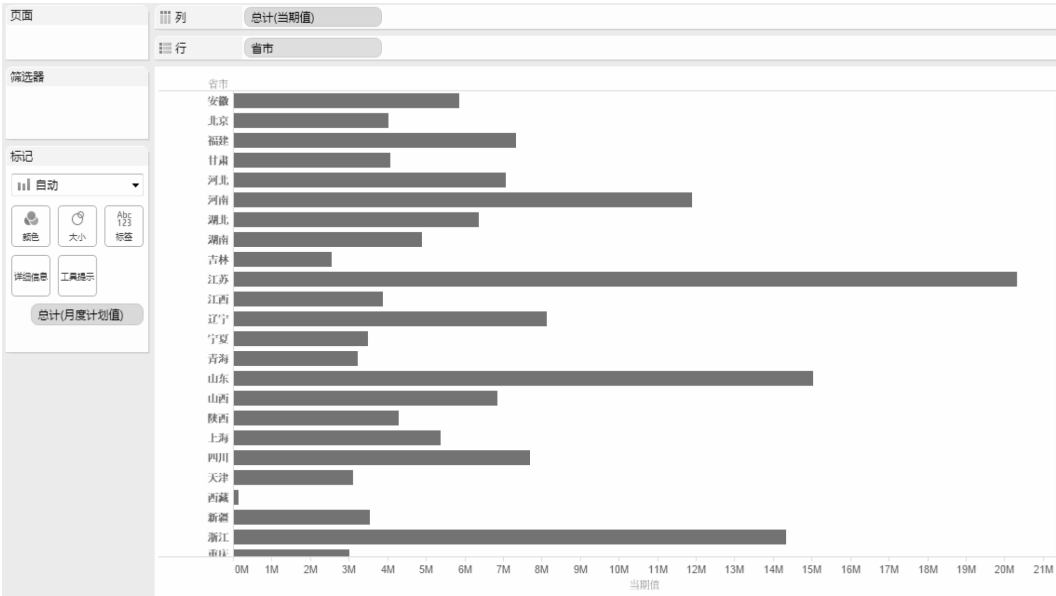


图4-43　创建标靶图时自动生成的条形图



图4-44　编辑参考线及参考区间

(3) 调整标记的大小后，得到标靶图，如图4-45所示。可以看出，2014年上半年，大部分省市售电量都已达到或超额完成计划标准，但河南和山东两省的计划完成情况较差，需进一步查询和分析原因。

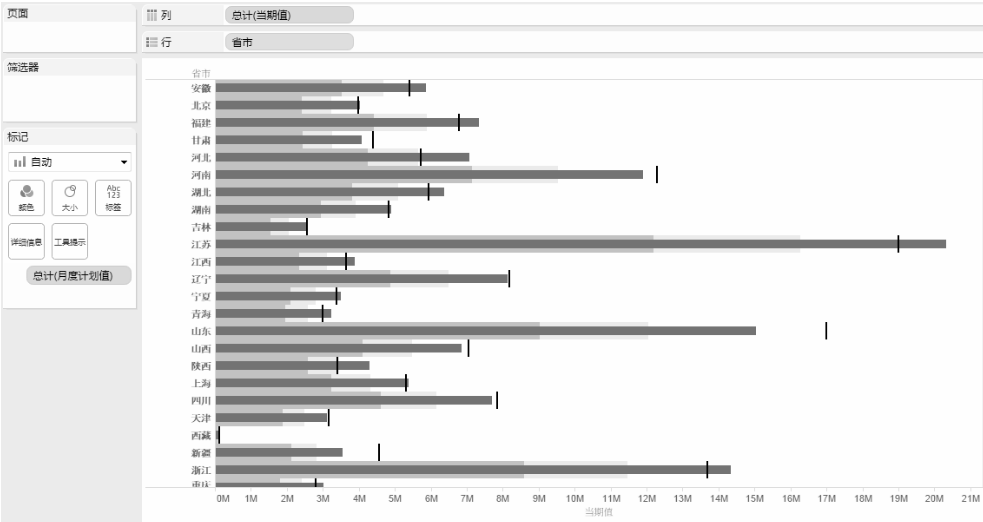


图4-45　添加参考线、参考区间后的标靶图

4.11　甘特图

甘特图，又称横道图，是以图示的方式通过活动列表和时间刻度形象地表示出任何特定项目的活动顺序和持续时间。甘特图的横轴表示时间，纵轴表示活动（项目），线条表示在整个期间上该活动或项目的持续时间，因此可以用来比较与日期相关的不同活动（项目）的持续时间长短。甘特图也常用于显示不同任务之间的依赖关系，并被普遍用于项目管理中。

下面以“比较各类物料不同供应商的延期交货情况”为例，说明创建甘特图的步骤和方法。

(1) 连接“某单位物资采购情况表”后，通过日期型字段“计划交货日期”和“实际交货日期”创建计算字段“延期天数”，如图4-46所示。



图4-46　创建“延期天数”

(2) 将“物资类别”和“供应商名称”拖放至行功能区，将“计划交货日期”拖至列功能区，并通过右键把日期级别更改为“日”。

(3) 将度量“延期天数”拖至“标记”卡上的“大小”后，生成图4-47所示的甘特图。



图4-47　甘特图

(4) 图4-47无法区分出各物资类别不同供应商的延期交货和提前交货情况。可将“延期天数”拖至“标记”卡的“颜色”上，并对其进行编辑，如图4-48所示。编辑颜色后的视图如图4-49所示。



图4-48　编辑“延期天数”的颜色（另见彩插图4-48）

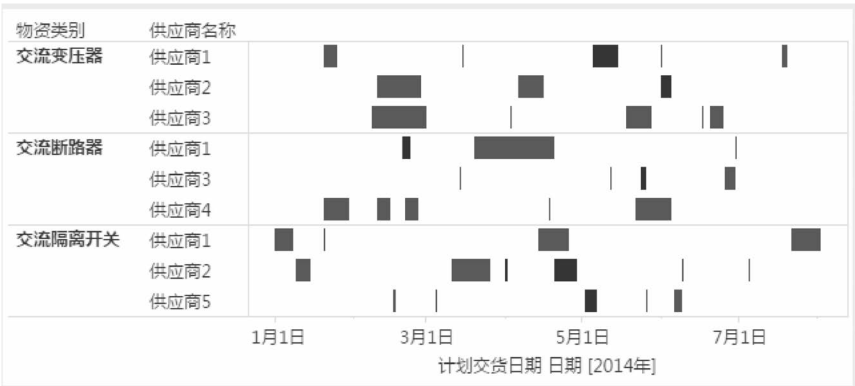


图4-49　供应商及时供货情况分析（另见彩插图4-49）

通过图4-49，可以分析比较三类物资的不同供应商的交付能力。对于“交流变压器”，供应商1的交付能力最好，供应商3延期情况最严重；而对于“交流断路器”，则是供应商3的交付能力最好；对于“交流隔离开关”，供应商5的交付能力最好，而供应商1的交付能力最差。该分析结果可对物资部采购人员制定未来采购计划提供帮助。

第 5 章

地图分析

Tableau地图功能十分强大，可实现省市、地市级的地图展示，并可编辑经纬度信息，实现对地理位置的定制化功能。本章主要介绍如何使用Tableau创建地图以及如何实现地图的高级功能。

其中，5.1节介绍创建基本地图、符号地图、填充地图以及混合地图的方法；5.2节介绍设置地理信息的方法，包括选择地图源和自定义地理编码等；5.3节介绍Tableau的高级地图功能和Tableau地图的高级应用，包括绘制多边形地图、创建背景图像视图和计算地理距离3项应用。

本章分析内容用到的数据源是2014年各省市售电量明细表、蒙东地理数据和分供电所线损率数据。其中售电量明细表包括每个地市售电量的当期值、累计值、同期值、同期累计值和计划值等字段；分供电所线损率数据包括该地市每个供电所的线损率字段。

5.1 地图简介

将Tableau连接到包含地理信息的数据源，并分配对应的“地理角色”后，Tableau可通过简单的拖放和单击生成地图。Tableau包含两种地图类型：符号地图和填充地图，同时也可制作包含两者的混合地图以及多维度地图。

本节重点介绍Tableau地图的基本概念和基本地图的创建步骤，通过本节的学习，读者将能够使用Tableau创建基本的地图视图。

5.1.1 分配地理角色

首先介绍Tableau对地理位置的识别。Tableau能够自动识别国家、省/直辖市、地市级别的地理信息，并能识别名称、拼音或缩写。

Tableau将每一级地理位置信息定义为“地理角色”，“地理角色”包括“国家/地区”“省/市/自治区”“城市”“区号”“CBSA/MSA”“国会选区”“县”“邮政编码”，其中只有“国家/地区”“省/市/自治区”“城市”对中国区域有效。具体地理角色定义如表5-1所示。

表5-1 Tableau地理角色定义

地理角色	说　　明
国家/地区	全球国家/地区，包括名称、FIPS 10、2 字符（ISO 3166-1）或3 字符（ISO 3166-1）。 示例：AF、CD、Japan、Australia、BH、AFG、UKR

（续）

地理角色	说　明
省/市/自治区	全世界的省/市/自治区，可识别名称和拼音。示例：河南、jiangsu、AB、Hesse
城市	全世界的城市名称，城市范围为人口超过1万、政府公开地理信息的城市，可识别中文、英文的城市名称。示例：大连、沈阳、Seattle、Bordeaux

一般情况下，Tableau会将“数据源”中包含地理信息的字段自动分配给相应的地理角色，此时，该字段在“维度窗口”中的字段图标为 ，表示Tableau已自动对该字段中的信息进行地理编码并将每个值与纬度、经度值进行关联，两个字段“纬度（生成）”和“经度（生成）”将自动添加到“度量窗口”，在创建地图时，可以拖放这两个字段进行展示。有时，Tableau会把地理信息字段识别为字符串字段，这种情况下，我们需要手动为其分配地理角色。可以在“维度窗口”中右键单击该字段，然后选择“地理角色”，为其分配对应的地理角色，之后该字段的图标将由 Abc 变为 ，如图5-1所示。

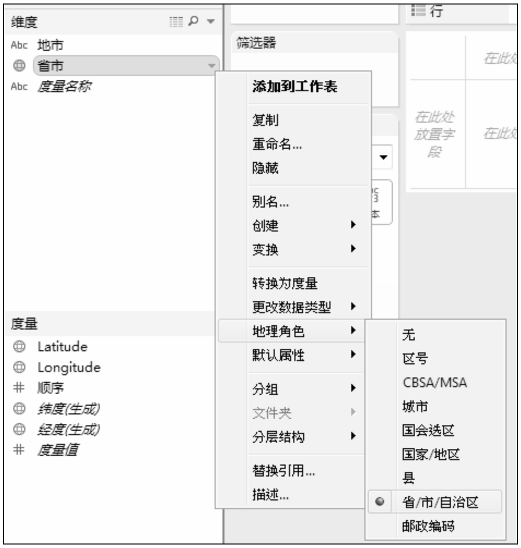


图5-1　地理角色分配

Tableau可以创建符号地图、填充地图、多维度地图和混合地图，下面将分别介绍这4种地图类型的视图创建步骤及创建后的效果。

5.1.2　创建符号地图

符号地图即以地图为背景，在对应的地理位置上以多种形状展示信息，图5-2即为一种符号地图。本节将以展示各省市某月售电量指标的各维度完成情况为例，介绍地图的创建方法。制作步骤如下。

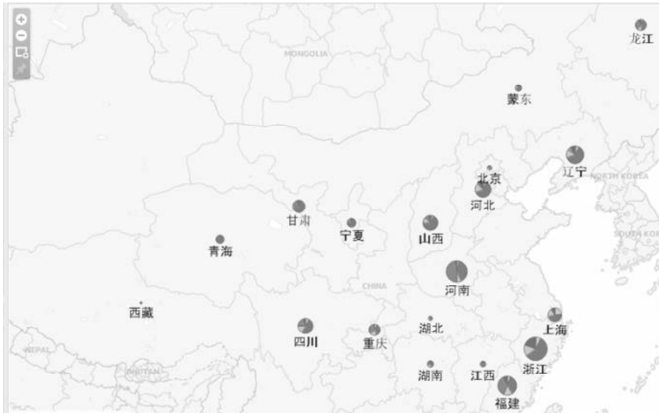


图5-2　符号地图

1. 创建符号地图

连接到数据并进行地理角色分配后，Tableau生成地图有以下两种方法。

方式1：双击“省市”字段，一般情况下，Tableau将自动调出地图视图，如未出现地图视图，则在“菜单栏”中选择“地图”➤“背景地图”，将“背景地图”设置为 “Tableau”即可。拖放“度量窗口”中“当期值”到“标记”卡中的“大小”生成符号地图，如图5-3所示。

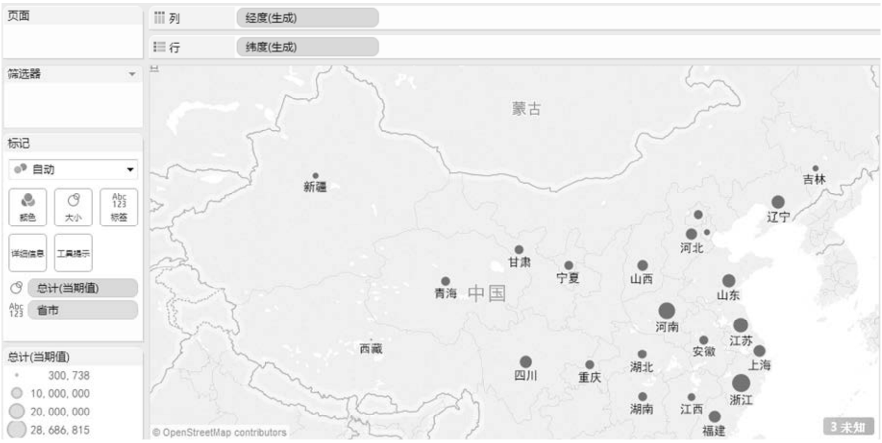


图5-3　生成符号地图

方式2：按住Ctrl，同时选中“维度窗口”中的“省市”和“度量窗口”中的“当期值”，单击“智能显示”，选中“符号地图”，Tableau默认将“当期值”作为“大小”在地图上进行展示，效果与方式1一致。

图5-3中的“圆”标记代表各个省市的售电量总量，如果想要分用电类别维度对比各个省市的售电量情况，只需把“标记”卡中图的类型由“圆”改为“饼图”，并把“用电类别”字段拖

放至“颜色”即可，如图5-4所示。

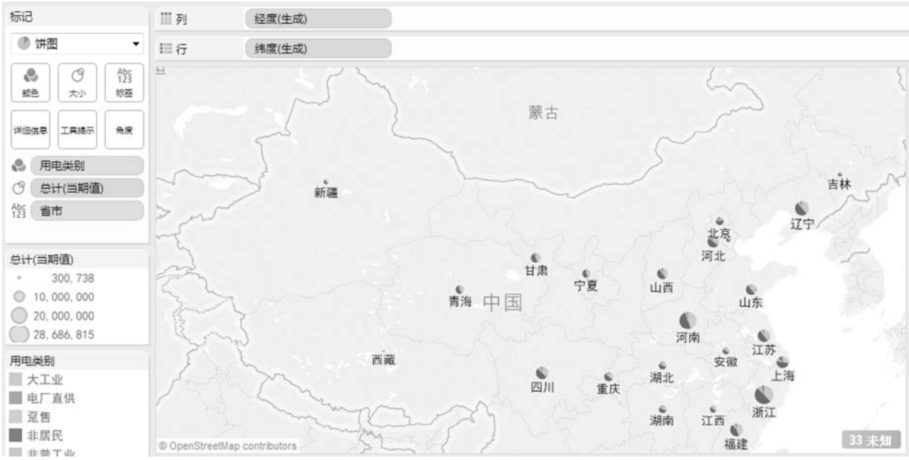


图5-4　更改形状为“饼图”

若要查看“地市”级别的信息，只需双击“维度窗口”中的“地市”字段，或拖动“地市”字段到“标记”卡中的“详细信息”即可，调整“标记”卡的相应内容后，生成的效果如图5-5所示。



图5-5 “地市”级别地图

2. 编辑地理位置

Tableau可对其地理库中不包含的地理位置进行信息编辑。

单击图5-3右下角的未知信息“3未知”，会弹出“[省市]的特殊值”对话框，在本对话框中有3个选项，单击“编辑位置”，或在“菜单栏”中选择“地图”➤“编辑位置”，此时弹出对话框，

如图5-6所示。

地理信息未识别有两种类型：①不明确，表示该数据所代表的地理位置有两个或以上，Tableau不知道该为其分配哪个位置；②无法识别，表示其不在Tableau的地理库中。

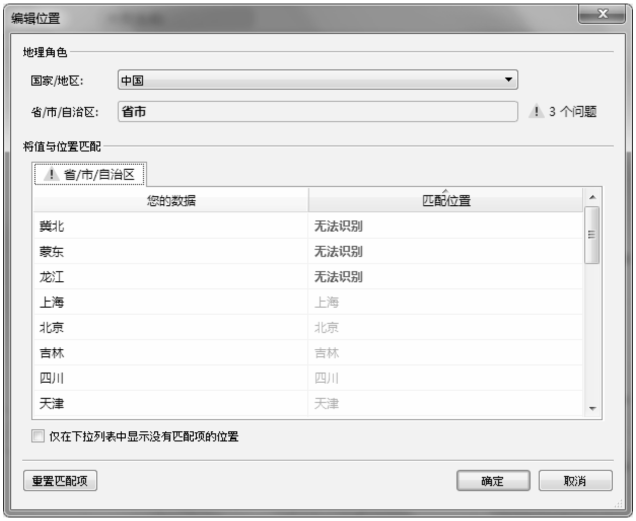


图5-6 “编辑位置”窗口

在“省/市/自治区”的标签页下，对“无法识别”的数据，可在“匹配位置”中选择一个“匹配项”，如“龙江”是Tableau无法识别的简称，可选择“黑龙江”将其映射到正确位置，如图5-7所示。

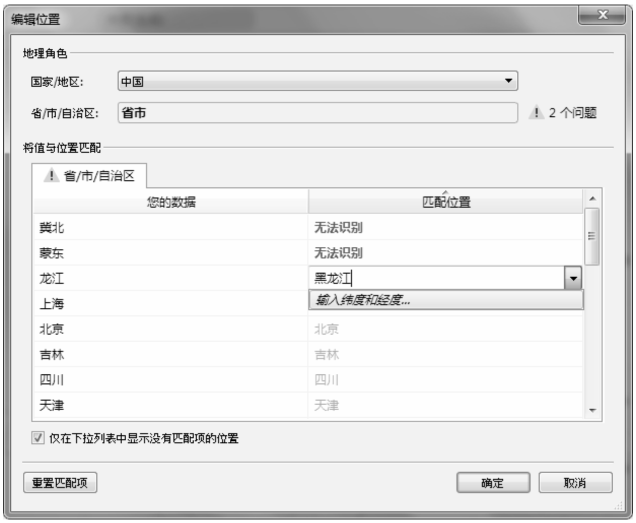


图5-7　将“龙江”映射为“黑龙江”

对“蒙东”和“冀北”的映射要从定位精度来考虑，若要绘制“省市”级别的地图，则需要对“蒙东”和“冀北”精确定位，在下拉列表中选择“输入纬度和经度”，在弹出的对话框中输入经纬度信息即可，如图5-8所示。若要绘制“地市”级别的地图，则只需将“冀北”映射到“河北”，“蒙东”映射到“内蒙古”即可，此时“冀北”和“内蒙古”包含的地市即可正确匹配。



图5-8　输入冀北的经纬度信息

说明　默认情况下，在匹配项下拉列表中会列出地理库中的所有位置，即中国的所有省市，选择“仅在下拉列表中显示没有匹配的位置”选项可限制列表只列出尚未匹配的位置，如图5-9所示。

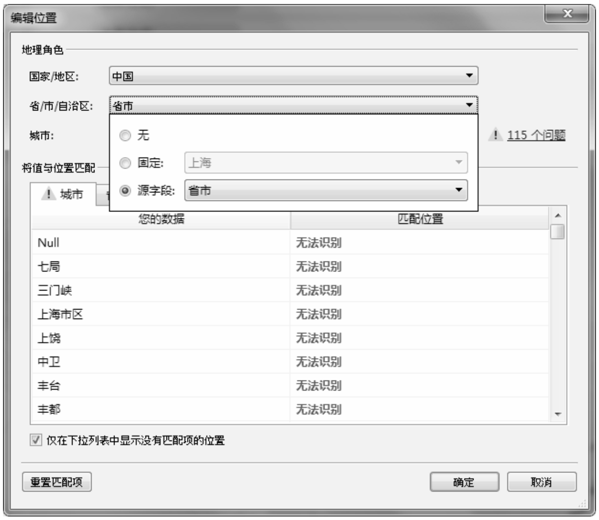


图5-9　“编辑位置”界面中选择位置

对地图中无法识别的“地市”位置，可采用相同的方式进行位置匹配。因Tableau的原理是先匹配上级单位，再匹配下级单位，所以对本案例来说，如果“冀北”“蒙东”“龙江”无法识别，其辖下的地市也会无法识别，因此一般情况下，需要先将上级地理角色的位置定义好，再设置下级地理角色。

若存在大量无法识别地理位置，逐个进行匹配或“输入纬度和经度”会耗费较大的工作量，

因此，建议通过“导入自定义地理编码”的方法，对Tableau的地理库进行扩充，实现地理位置识别，具体操作可参考5.2.2节。

说明　单击地图右下角的未识别信息，弹出的对话框中有3个选项，“编辑位置”已介绍使用方法，剩下的“筛选数据”和“在默认的位置显示数据”主要是针对数据信息的设置，“筛选数据”为将不识别的数据剔除，“在默认位置显示数据”是将经纬度设置为0进行展示。因此，在地图模式下，不建议选择。

3. 设置地图格式

生成地图后，有多个选项可设置地图的显示效果。调用地图格式设置的路径为：选择“地图”➤“地图选项”打开地图选项窗口。

在地图选项中，有“后台”“地图层”和“数据层”3个选项卡可供设置，其中“数据层”为Tableau预设的美国人口普查等信息，对中国情况不适用，因此不作详细介绍。下面我们主要介绍后台和地图层的使用。

后台：使用地图选项窗口顶部的“样式”下拉菜单，在普通、灰色或黑色地图样式之间进行选择（见图5-10），可设置不同的地图显示风格（见图5-11）。



图5-10　地图选项



图5-11　不同地图样式的展示效果

除地图样式选项之外，还可以使用“冲蚀”滑块控制背景地图的颜色强度或亮度。滑块向右移得越远，地图背景颜色就越浅。冲蚀效果如图5-12所示。

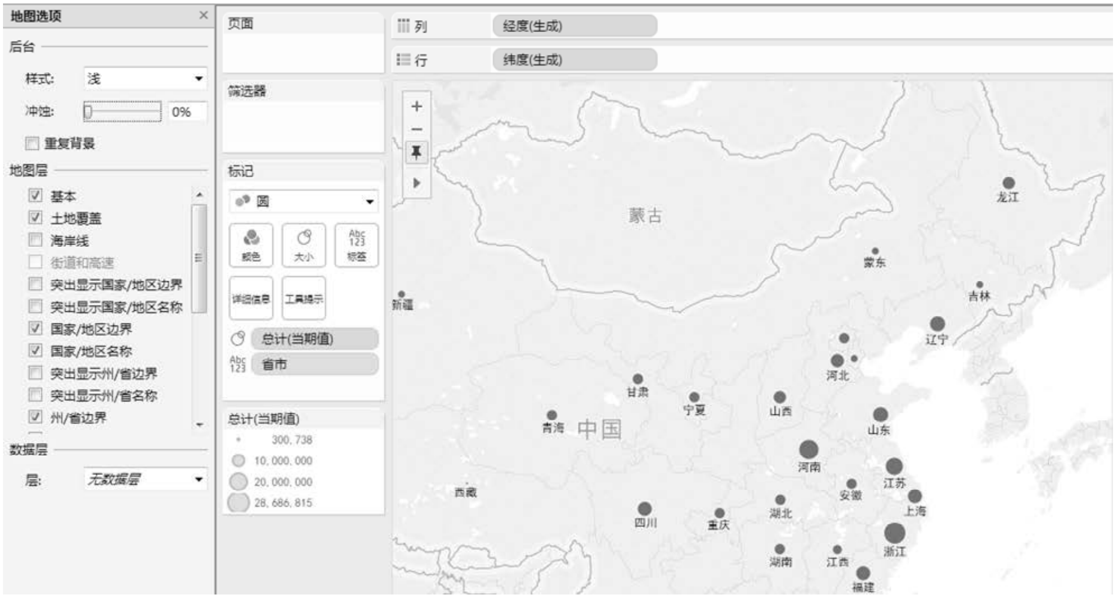


图5-12　冲蚀设置效果

地图层：Tableau 地图提供了多个层，这些层可对地图上的相关点进行标记。Tableau提供的部分地图层仅在特定缩放级别上可见。表5-2列出了每个地图层及其使用的范围。

表5-2　Tableau 地图层说明

层　名　称	说　　明
基本	显示包括水域和陆域的底图
土地覆盖	遮盖自然保护区和公园以便为地图提供更大深度
海岸线	以深灰色显示海岸线轮廓
街道和高速	标记公路、高速公路以及城市街道，还包括公路和街道名称
突出显示国家/地区边界	以浅灰色显示国家/地区轮廓
突出显示国家/地区名称	以浅灰色显示国家/地区名称
国家/地区边界	以深灰色突出显示国家/地区边界
国家/地区名称	以深灰色突出显示国家/地区名称
突出显示州/省边界	以浅灰色显示州/省边界
突出显示州/省名称	以浅灰色显示州/省名称
州/省边界	以深灰色突出显示州/省边界
州/省名称	以深灰色突出显示州/省名称

地图层有不同的设置效果，如图5-13和图5-14所示。



图5-13 地图层不同的设置效果1



图5-14 地图层不同的设置效果2

说明 在展示世界地图时，可选择“重复背景”，此时背景地图可多次显示相同区域。在非“重复背景”下，世界地图只展示一次。

单击地图选项窗口底部的“设置为默认值”，即将设置好的地图格式设置为默认值，此时，在本Tableau中创建的地图均采用本次设置。

5.1.3 创建填充地图

填充地图即将地理信息作为面积进行填充，图5-15即为一种填充地图。创建填充地图的方法有以下3种。



图5-15　填充地图（另见彩插图5-15）

- 双击“维度窗口”中的“省市”字段，生成符号地图后，拖放度量窗口中“当期值”到“标记”卡中的“颜色”。
- 按住Ctrl，同时选中“维度窗口”中的“省市”和“度量窗口”中的“当期值”，单击“智能显示”，选中“填充地图”，Tableau默认将“当期值”作为“颜色”在地图上进行展示。
- 创建好符号地图后，在“标记”卡的图形选项中选择“已填充地图”或在“智能显示”中选择“填充地图”。

这3种方法创建的填充地图效果一致，均如图5-15所示。

在填充地图中，对Tableau不能识别的位置，无法通过编辑位置来实现地理定位。如要实现这部分地理位置的填充，请参考5.3.1节多边形地图。

填充地图只能识别到“省/市/自治区”，不能识别“城市”一级的地理角色（如图5-16所示），因此若要展示城市信息，只能采用符号地图。



图5-16　填充地图不可识别城市

5.1.4　创建多维度地图

多维度地图通过对不同维度的信息用多个地图展示，实现信息的分维度比对。多维度地图展示要在已创建好的符号地图或填充地图的基础上创建。以符号地图或填充地图为基础创建多维度

地图的步骤是相同的，本节将以填充地图为例，介绍如何创建多维度地图。

在上一节创建好的填充地图基础上，拖放“维度窗口”中的“统计周期”到行功能区，“用电类别”到列功能区，并放在“经度（生成）”“维度（生成）”之前，即可实现分时间段、分用电类型的各省市的指标完成值的对比分析，如图5-17所示。

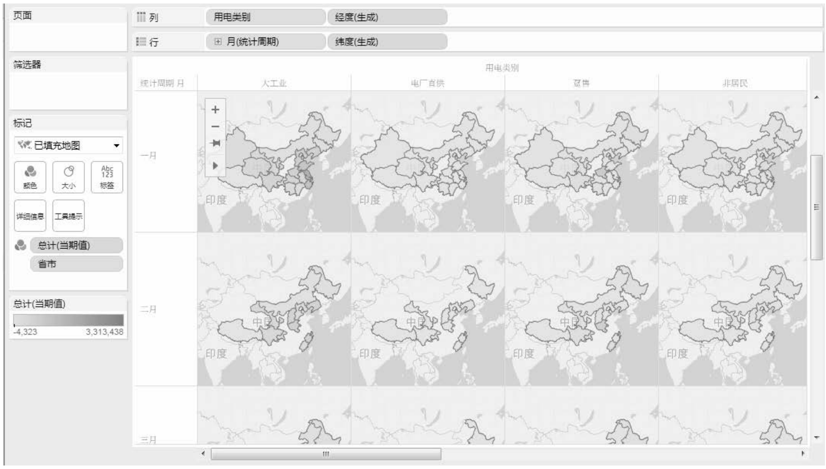


图5-17　多维度地图

对行标题、列标题进行格式设置后，效果如图5-18所示。

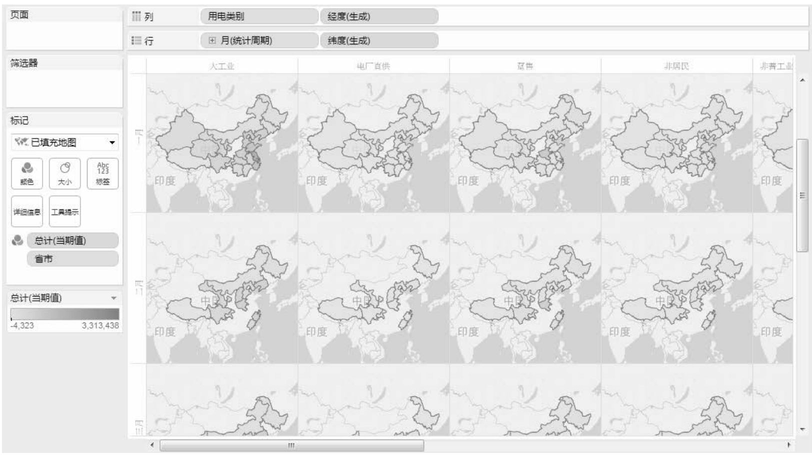


图5-18　完成格式设置的多维度地图

5.1.5　创建混合地图

混合地图是指把符号地图和填充地图叠加而形成的一种地图形式，如图5-19所示。

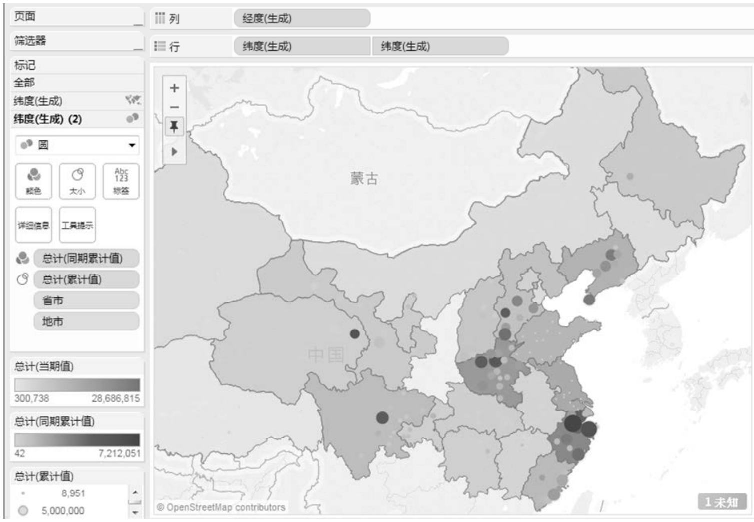


图5-19　省市和地市的混合地图

创建混合地图时，首先需要创建一个符号地图或填充地图，下面以先创建填充地图为例。首先生成以“当期值”为颜色的分省市填充地图，然后将“度量窗口”中的“纬度（生成）”再次拖放到行功能区上，或在行功能区中，按住Ctrl键拖放“纬度（生成）”到右侧，此时同时展示两个地图，如图5-20所示。右键单击行功能区的“纬度（生成）”，选择“双轴”，两个地图重叠为一个，生成如图5-21所示的效果。同理重复拖放列功能区的“经度（生成）”，也可实现同样的效果。



图5-20　生成两个地图

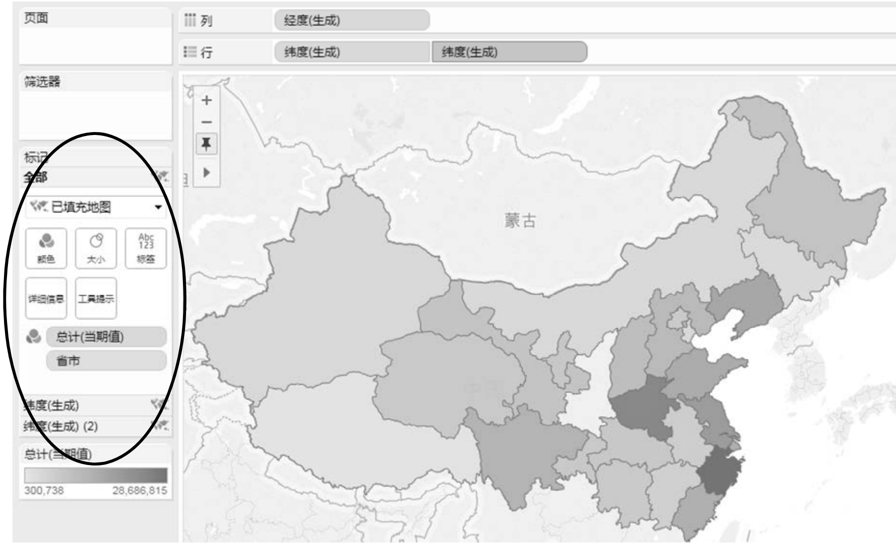


图5-21　“双轴”设置后的地图

可以看到在“标记”卡中生成了两个切换条，为“纬度（生成）”和“纬度（生成）（2）”，分别代表两个地图图层。

选择“纬度（生成）（2）”，修改其图形类型为“圆”，拖放“度量窗口”中“同期累计值”字段到“标记”卡中的“颜色”，“累计值”到“大小”，编辑“颜色”为红色系，拖放“维度窗口”中的“省市”到“标签”，生成如图5-22所示的效果，实现了三维信息的展示。

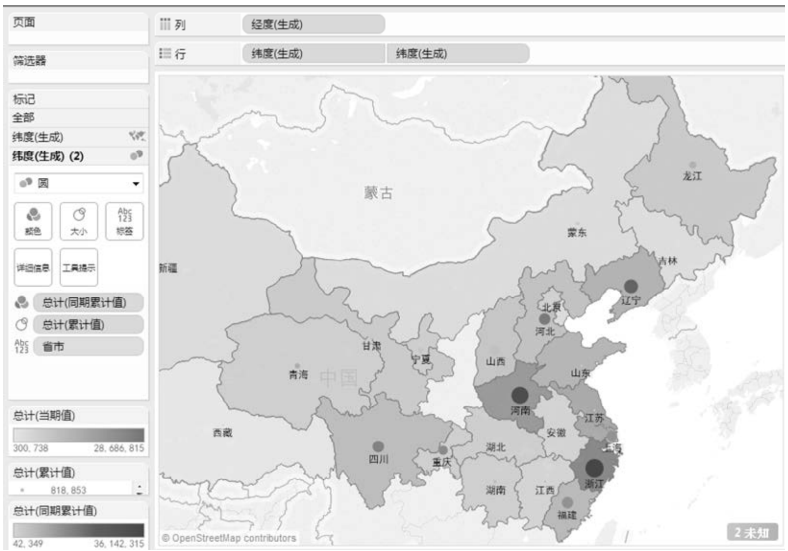


图5-22　生成混合地图

若要展示地市级信息，则在“维度（生成）（2）”这一地图层双击“度量窗口”中的“地市”，或拖放“地市”到“标记”卡中的“详细信息”，并拖放“地市”到“标签”中，最终生成如图5-19所示的混合地图。

说明　在创建混合地图时，创建两个地图图层后，先设置每个地图图层展示的信息，再设置“双轴”显示，实现效果一致。

5.2　设置地理信息

本节将介绍如何设置和使用Tableau的地图源，并对如何通过自定义地理编码扩充Tableau的地理信息库进行说明。

5.2.1　选择地图源

Tableau中的背景地图选项为使用者提供了地图源的多种选择，用户可以选择不使用地图源，或选择Tableau自带的地图源“Tableau”，或脱机使用地图，或使用WMS服务器实现自定义地图源“WMS服务器（W）...”，并可设置何种地图源为默认地图源，如图5-23所示。



图5-23　菜单栏中“背景地图”调用位置

1. 联机地图

默认情况下，所有新建工作表都会自动连接到Tableau的联机地图源“Tableau”，其地理位置信息由开源地图供应商OpenStreetMap提供。

若用户习惯使用某地图源，可将其指定为Tableau默认地图源，操作方式为在“地图”➤“背景地图”菜单中选择地图源，然后选择“地图”➤“背景地图”➤“设置为默认值”。

2. 地图存储和脱机工作

在使用联机地图创建地图视图时，Tableau 会将构成地图的图像存储在缓存中。这样在进行分析时，就不必等待检索地图。同时，通过存储地图，可以在设备脱机时仍使用部分地图进行分析。地图的缓存将随Internet Explorer 的Internet 文件一起存储，删除Internet Explorer中的临时文件即清除了地图缓存。

在脱机工作并打开地图视图时，将自动使用存储的图像。若之前未缓存地图视图，在“地图”菜单中选中“背景地图”➤“脱机”，Tableau也可展示国家级别的地图。不过，当需要检索新的地图时，如果缓存中无该图像，则在联机之前将无法加载该地图。下面列出了要检索新的地图的主要操作。

- 打开层：如果打开未存储在缓存中的层，例如在地图选项中，设置打开“州/省名称”等地图层，则Tableau 需要进行网络连接以检索所需信息。
- 缩放：放大或缩小地图需要不同的地图图像，如果缓存中不存在指定缩放级别上的图像，则Tableau 需要检索新的地图。
- 平移：如果脱机工作并且未将所需地图图像存储在缓存中，则不会加载新图像。

为防止地图图像信息过时，存储的地图图像在30天内有效，之后Tableau 将不使用存储的图像，而是要求重新连接并获取更新的地图。

3. WMS服务器

如果具有提供特定行业的WMS 服务器，Tableau可以添加该服务器作为地图源。在添加了WMS 地图服务器之后，可以导出地图源与他人共享，或导入共享的地图源。

(1) 添加WMS 服务器：选择“地图”➤“背景地图”➤“WMS服务器（W）...”，单击“添加”。在随后显示的对话框中，键入该服务器的URL，然后单击“确定”，如图5-24所示。



图5-24　添加WMS服务器

Tableau可以添加若干地图服务器。每个地图源都显示在“WMS服务器”菜单上的地图源列表中，地图源将随工作簿一起保存，并且可供共享该工作簿的任何人使用。

(2)导出WMS 服务器：选择“地图”➤“背景地图”➤“WMS服务器（W）...”，选择要作为地图源导出的服务器，然后单击“导出”，键入文件名并选择位置，该文件会另存为Tableau 地图源（.tms），单击“保存”，如图5-25所示。

导出的Tableau 地图源包括已设置好的任何“地图选项”内容，如已指定为默认显示的地图层。如果更改“地图选项”设置，则应再次导出该地图源以更新设置。

(3) 导入地图源：选择“地图”➤“背景地图”➤“WMS服务器（W）...”，单击“导入”，导航到保存的地图源文件（.tms），选择该文件，然后单击“打开”。

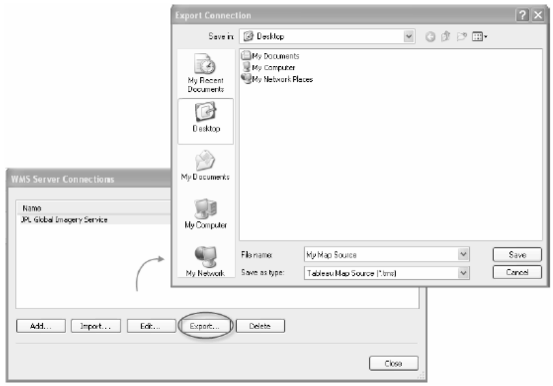


图5-25　导出WMS服务器

5.2.2　自定义地理编码

当有大量Tableau无法识别的地理位置时，可通过导入自定义地理编码扩充Tableau的地理信息库。注意，自定义地理编码只能绘制符号地图。

本节将介绍如何导入和管理自定义地理编码，以及自定义地理编码的数据格式要求。

1. 准备自定义地理编码文件

在 Tableau 中添加自定义地理编码时，需要导入包含新地理角色及其纬度/经度定义的文件，这些文件应采用.CSV格式。有以下3种类型的自定义地理编码。

- **扩展现有角色**

这是对现有地理角色的信息补充，例如，Tableau默认只能识别人口1万以上且政府公开地理信息的城市，因此，内置的城市并不完整，通过扩展现有角色，可以扩充可展示的城市范围。

自定义地理编码文件中，列的名称定义了地理角色，如果想扩展地理角色，则必须确保列名称与Tableau中现有地理角色的名称一致，同时要包括其所有上级地理角色。以要扩展“城市”这个地理角色为例，地理编码文件应与表5-3中的数据结构保持一致。

表5-3　扩展现有角色示例

国家/地区	省/市/自治区	城市	经度	纬度
中国	辽宁	朝阳（辽）	120.45	41.57
中国	北京	朝阳（京）	116.43	39.92
中国	重庆	重庆市区	106.55	29.57
中国	重庆	南岸	106.57	29.52
中国	重庆	江北	106.57	29.6
中国	重庆	北碚	106.4	29.8

说明 在自定义地理编码文件中，不能出现重复的经度、纬度信息，否则Tableau无法区分不同的地理位置，会导致无法导入文件。

- **添加新角色**

除了扩展已有地理角色外，Tableau还可以添加新的地理角色。例如，需要在地图上绘制变电站的位置，并要在“国家/地区”➤“省/市/自治区”➤“城市”的现有分层结构之下添加。

此类地理编码的导入文件必须包含现有分层结构中的每级地理角色，这样才能在各级地理角色之间构建关联关系。例如，变电站代码将添加到“国家/地区”➤“省/市/自治区”➤“城市”分层结构中，导入文件必须包含“国家/地区”“省/市/自治区”和“城市”级别的所有列。表5-4是包含变电站经纬度信息的文件示例。

表5-4 添加新地理角色示例

国家/地区	省/市/自治区	城市	变电站	经度	纬度
中国	重庆	南岸	变电站A	106.56	29.52
中国	重庆	南岸	变电站B	106.57	29.54
中国	重庆	南岸	变电站C	106.55	29.57

导入“变电站”编码文件后，会在分配“地理角色”中出现“变电站”选项，如图5-26所示。



图5-26 添加新的地理角色“变电站”的效果

- **添加新分层结构**

在扩展现有角色或添加新角色时，是在现有分层结构内进行操作，若需要添加全新分层结构，例如需要按“区域”“行政省市”自定义地理分层结构，这种情况类似于创建新角色，区别是没

有映射到现有父角色的列。这时需要创建多个导入文件，每个文件表示新分层结构中的一个级别，示例见表5-5。

表5-5　添加新分层结构的多个文件示例

区域	经度	纬度	区域	行政省市	经度	纬度
华东	119.51	32.25	华北	冀北	116.57	41.24
华北	115.13	38.17	华北	河北	115.07	37.97
西北	100.57	34.13	华北	山东	118.88	36.02
华中	113.01	33.98	华北	天津	117.19	39.29
东北	125.69	45.07	……	……	……	……

在自定义地理编码文件中确保每个文件都包含当前级别和其上面所有级别的列。将每个导入文件保存到一个文件夹中，并使用文件所表示的级别来命名每个文件。对于此示例，有名为“区域”“行政省市”的两个导入文件。

导入文件后，会在“地理角色”中出现“区域”“行政省市”选项，如图5-27所示。



图5-27　添加新的分层结构“区域”　“行政省市”的效果

说明　自定义地理编码文件中的经度和纬度必须是实数而不是整数，即在编辑这些值时，要确保包含至少一个小数位。

2. 导入自定义地理编码文件

准备好导入文件之后，可以将自定义地理编码导入到 Tableau 中。操作步骤为：选择“地图”➤“地理编码”➤“导入自定义地理编码（I）...”，在“导入自定义地理编码（I）...”对话框中，输入包含导入文件的路径，单击“导入”，Tableau会扫描所选目录中的所有 CSV 文件并替换已导入的地理编码，新地理角色变为可用，然后可将数据中的地理数据分配到导入的地理角色中。

3. 管理自定义地理编码文件

在自定义地理编码的文件进行更新后，选择“地图”➤“地理编码”➤“刷新打包的地理编

码（S）”即可将Tableau中已导入的地理编码进行更新。

在将工作簿保存为打包工作簿时，自定义地理编码数据会随工作簿一起打包。

通过选择“地图”➤“地理编码”➤“移除自定义地理编码（R）”即可移除之前导入的自定义地理编码文件。

5.3 高级功能

通过前两节的学习，我们已经可以利用Tableau绘制普通的地图，并可进行简单的地理信息设置。本节将重点介绍如何利用Tableau实现地图的高级功能，包括高级填充地图——多边形地图，高级符号地图——背景图像，以及利用地图功能计算地理距离等。

5.3.1 多边形地图

多边形地图是对填充地图的一种补充，基于地理编码数据文件绘制一个多边形区域，实现自定义填充地图。本节将以绘制蒙东电力公司的供电区域为例来介绍操作方法。

1. 准备地理信息数据

在创建自定义区域时，首先应准备一份要绘制区域边界各点的经纬度和绘制顺序的文件，数据结构如表5-6所示，其中“顺序”为Tableau连接区域边界各点的绘制顺序，Latitude为区域边界各点的纬度，Longitude为区域边界各点的经度。Tableau会按照顺序将各点经纬度信息连接成线，生成填充地图。

表5-6 绘制“蒙东”的数据结构

省市	地市	顺序	Latitude	Longitude
蒙东	通辽市	1	45.2911	119.2346
蒙东	通辽市	2	45.4724	119.3225
蒙东	通辽市	3	45.5328	119.4434
蒙东	通辽市	4	45.6152	119.5093
蒙东	通辽市	5	45.6152	119.5532
……	……	……	……	……

蒙东供电公司包括呼伦贝尔、兴安盟、赤峰市、通辽市4个地市，在地理信息数据中应包括这4个地市边界各点的经纬度信息，共计718个点。数据存储为《蒙东地理数据》。

2. 创建视图

下面以构建蒙东分地市的售电量完成情况为例介绍如何使用多边形地图，创建步骤如下。

(1) 连接“2014年各省市售电量明细表数据”和“蒙东地理数据”两张表，并以“省市”“地市”为联接建立内部联接模式，如图5-28所示。



图5-28　建立数据联接关系

(2) 转到工作表后，双击Latitude和Longitude，调出地图页面。

(3) 拖放“地理数据”中的“地市”到“详细信息”，修改“标记”卡中图形展示为“多边形”。

(4) 拖放“顺序”到“标记”卡中的“路径”，生成图5-29所示的效果。

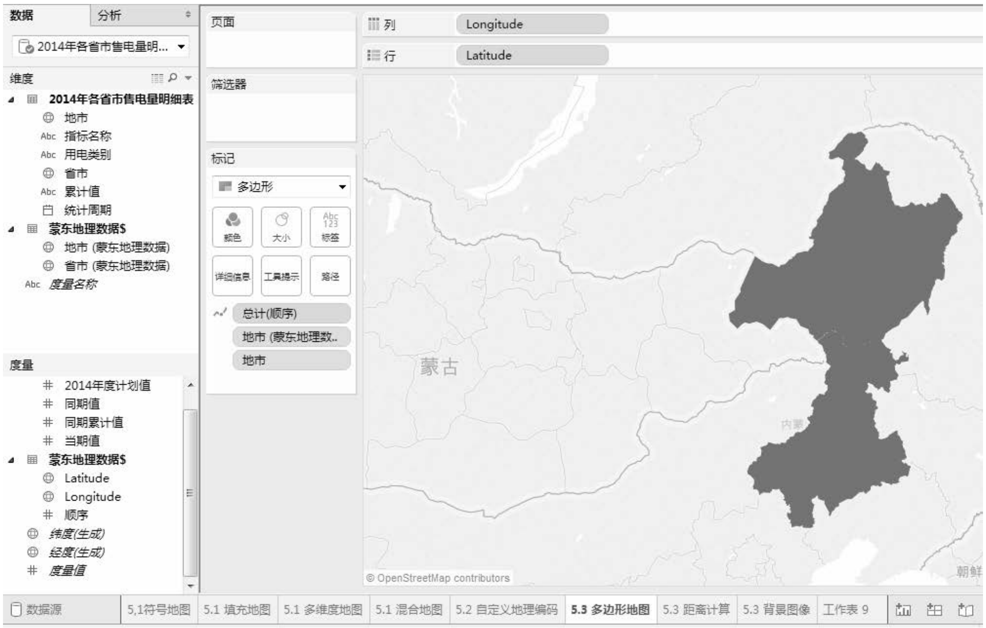


图5-29　创建的蒙东地图

(5) 拖放“当期值”到“标记”卡中的“颜色”，调整“标记”卡中“颜色”“工具提示”等设置，最终效果如图5-30所示。



图5-30　蒙东某指标的分地市完成情况

5.3.2　背景图像地图

Tableau可以添加图片作为数据的背景图像，可通过自定义图像更换地图的展示形式，为数据提供更多位置信息；还可对Tableau不支持的地图细度进行扩展，展示更为详细的地图信息等。

本节将以创建某地市的分供电所线损率对比图为例介绍背景图像视图的创建方法。

1. 导入背景图像

添加背景图像的步骤如下。

(1) 右键单击要添加的图片，查看属性，在“详细信息”中查看图片的宽度和高度，如图5-31所示。

(2) 在分供电所线损率数据中新增两个字段X、Y，并新增一行数据，其中X字段为图片的宽度2380，Y字段为图片的高度1408，将数据导入Tableau，采用“实时”连联接方式。

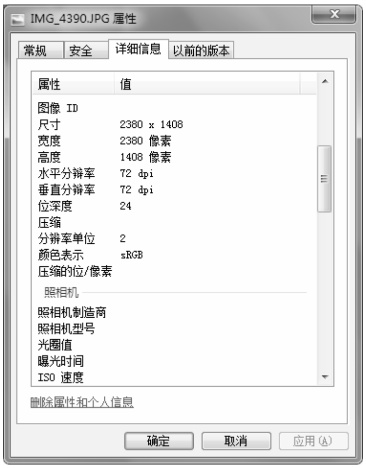


图5-31　查看图像属性

(3) 在Tableau菜单栏中，选择“地图”➤“背景图像”，选择数据源，单击“添加图像”，在“编辑背景图片”对话框中，单击“浏览”选择要添加到Tableau中的图像，也可以键入URL以链接到在线提供的图像，并可在名称栏修改地图名称。将数据中的X字段映射到图像X轴，Y字段映射到图像的Y轴，设定X字段的右边最大值为图像宽度2380，Y字段的上边最大值为图像高度1408，建立背景图像的X-Y坐标系，单击“确定”，如图5-32所示。



图5-32　添加背景图片

(4) 将“度量窗口”中X、Y字段分别拖放到列功能区和行功能区，此时背景图像已导入，如图5-33所示。

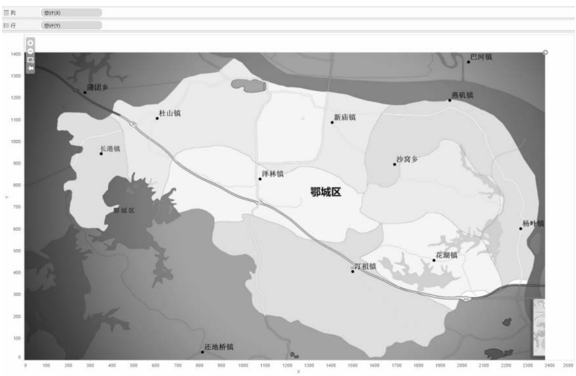


图5-33　背景图像导入

在背景图像添加进来后，随时可对其进行编辑，操作步骤为：选择“地图”➤“背景图像”，在“背景图像”对话框中，选择要编辑的图像，然后单击“编辑”（也可以直接双击图像名称），在“编辑背景图片”对话框中可针对图像的导入设置进行调整，也可使用“选项”功能进行设置。其中，“锁定纵横比”即适中保持图像的原始比例，取消选择此选项则允许图像变形；“始终显示

整个图像”即禁止剪裁图像，如果在视图中将两个轴都锁定，则此选项可能无效；“添加图像显示条件”在添加多个背景图像时适用，为避免在工作表上多图像重叠显示，可以设置“仅在以下时间显示”某个背景图片，添加筛选字段，定义仅在某种情况下显示该背景图片，如图5-34所示。



图5-34　背景图片编辑功能

说明　Tableau支持添加多个背景图像，可以选择在每张工作表上激活的图像，即针对一组工作表使用一个图像集。使用“背景图像”对话框中的复选框可以为当前工作表启用或禁用图像。启用多个图像，并设置合理的X、Y的位置，可以在一个工作表中显示多个图像。

2. 准备展示数据

在添加完背景图片后，需要定位各信息点的X轴、Y轴的坐标，定位方法为：在背景图片中，选中需展示信息的某点，右键单击，在弹出对话框中单击“添加注释”，选择注释方式为“点”，此时弹出如图5-35所示的对话框，单击“确定”，此时展示效果如图5-36所示，对注释所指的点的位置可随意拖动进行编辑，X、Y的坐标会随着位置的变化而变化。



图5-35　添加注释定位坐标



图5-36　添加注释后的界面展示

根据注释中X、Y所指的坐标信息，在源数据中为每个供电所的X、Y两列中添加坐标信息，添加完的效果如表5-7所示。

表5-7　为源数据添加X、Y字段信息

供电所	线损率	X	Y
长港	0.03	343.8	940.73
杜山	4.35	608.55	1101.69
华容	8.31	1034.27	1257.28
泽林	4.87	1080.32	823.04
新庙	1.99	1404.54	1082.02
汀祖	2.53	1492.19	402.4
花湖	3.91	1882.144	454.27
沙窝	5.13	1907.18	803.02
燕矶	7.71	1966.21	1105.27
杨叶	2.37	2266.73	600.92

3. 构建视图

在准备好数据后，在Tableau的数据源中刷新数据，拖动“供电所”字段到“标记”卡的“详细信息”，即可按照常规创建符号地图的方式创建地图。本例中以线损率为大小在背景图片上做展示，整体效果如图5-37所示。

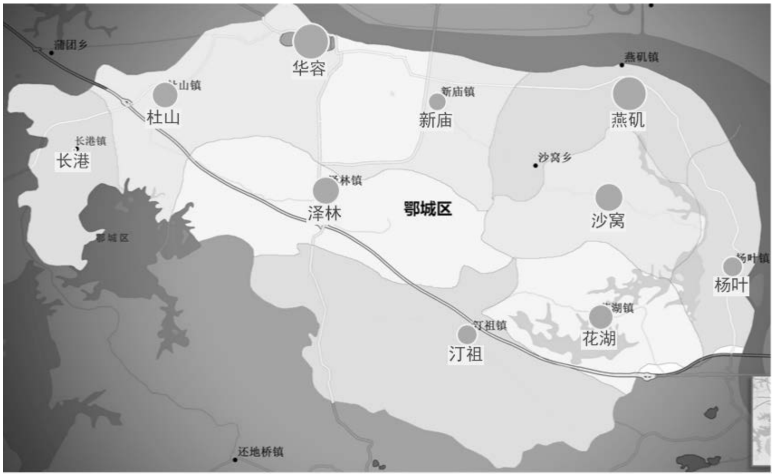


图5-37　背景图像展示效果

5.3.3　地理位置距离计算

Tableau创建地图后，可利用自带的公式功能计算地图上两点之间的距离，在实际应用中，可通过距离计算功能来计算变电站之间的距离，从而粗略测算不同布线方式的优劣。

本节利用5.2.2节中自定义地理编码创建的3个变电站的地理数据进行演示，创建步骤如下。

(1) 复制5.2.2节中表5-4的变电站数据，使每个变电站都存在两行数据，并添加3列数据，分别为表1变电站、表1Lon、表1Lat，代表变电站名称、经度、纬度，并让每个变电站分别对应另外两个变电站，生成如表5-8所示的文件。

表5-8　变电站数据

国家/地区	省/市/自治区	城市	变电站	Longitude	Latitude	表1变电站	表1Lon	表1Lat
中国	重庆	南岸	变电站A	106.56	29.52	变电站B	106.57	29.54
中国	重庆	南岸	变电站A	106.56	29.52	变电站C	106.55	29.57
中国	重庆	南岸	变电站B	106.57	29.54	变电站A	106.56	29.52
中国	重庆	南岸	变电站B	106.57	29.54	变电站C	106.55	29.57
中国	重庆	南岸	变电站C	106.55	29.57	变电站A	106.56	29.52
中国	重庆	南岸	变电站C	106.55	29.57	变电站B	106.57	29.54

(2) 导入创建好的数据表，并创建新的字段“距离计算”，计算公式为：`3959*ACOS(SIN(RADIANS([Latitude]))*SIN(RADIANS([表1Lat]))+COS(RADIANS([Latitude]))* COS(RADIANS([表1Lat]))*COS(RADIANS([表1Lon])-RADIANS([Longitude])))`，如图5-38所示。

(3) 双击度量窗口的Latitude和Longitude，生成3个变电站的地图视图，改变“标记”卡中的图形展示为“线”。

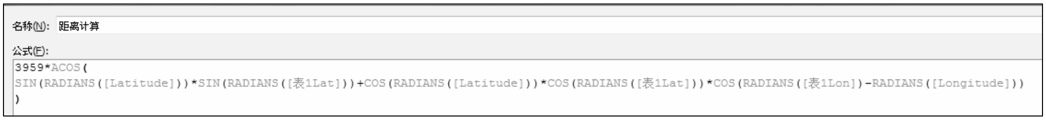


图5-38　构建计算字段

(4) 拖动“距离计算”字段，到“标记”卡中的“颜色”或“大小”，右键单击“标记”卡中的“距离计算”字段，选择“维度”，如图5-39所示。



图5-39　设置汇总方式为“维度”

(5) 调整“标记”卡中的“颜色”“大小”和“标签”设置，最终生成效果如图5-40所示。可知变电站A、B、C之间的距离分别为3507 m、1507 m、2396 m。

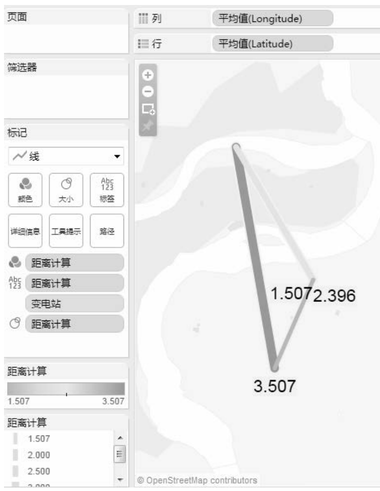


图5-40　距离计算效果

第 6 章

高级数据操作

本章主要介绍如何创建分层结构、组、集、参数、计算字段、参考线与参考区间，以及如何灵活运用它们来创建视图。熟悉和掌握Tableau的这些高级数据操作方法将有助于了解Tableau的数据组织形式和基本工作方式，它是进行高级可视化分析的基础，建议在阅读第7章之前，首先掌握本章介绍的Tableau高级数据操作。

6.1节和6.2节描述了分层结构和组，两者经常结合使用以实现对数据的上钻和下钻操作；6.3节描述了集的创建与使用方法；6.4节描述了参数的创建和使用方法；6.5节介绍了如何创建和使用计算字段，以及表计算、详细级别表达式、百分比这3种函数的使用方法；6.6节介绍了参考线以及参考区间在工作表视图中的使用方法，有效地丰富了所创建的视图。

本章所用到的数据为坐席接听统计数据，其中每条记录包括一个坐席每天接听电话的统计信息，以及该坐席所属中心、部门、组、班、工号、姓名等基本信息。

6.1　分层结构

分层结构（hierarchy）是一种维度之间自上而下的组织形式。Tableau默认包含了对某些字段的分层结构，比如日期、日期/时间、地理角色。以日期维度为例，日期字段本身包含了“年–季度–月–日”的分层结构。

除了Tableau默认内置的分层结构外，针对多维数据源，由于其本身包含了维度的分层结构，所以Tableau直接使用数据源的分层结构。针对关系数据源，Tableau允许用户针对维度字段自定义分层结构，在创建分层结构后，将显示在维度窗口中，其字段图标为 品 。

分层结构对维度之间的重新组合有重要作用，上钻（drill up/roll up）和下钻（drill down）是导航分层结构的最有效方法。例如，在查看不同月份的人工服务接听量时，单击列功能区上的⊞控件，可以下钻查看每月各日的接听量；单击列功能区上的⊟控件，上钻查看每月的接听量，如图6-1所示。

下面我们介绍如何在关系数据源中创建和使用分层结构。

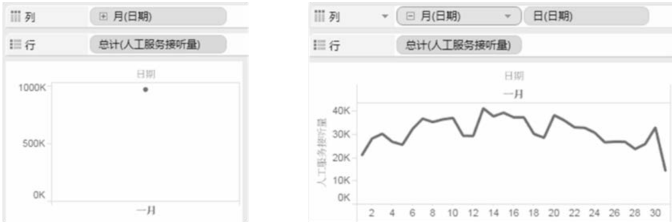


图6-1　人工服务接听量折线图

6.1.1　创建分层结构

图6-1展示的是一个月人工服务接听量的折线图，如果希望查看不同中心以及下级各个部门、各个组的人工服务接听量，依据已有的维度字段“中心”“部”和“组”来创建分层结构即可轻松实现。

1. 方式1：通过拖动方式创建名为“组织”的分层结构

在“维度窗口”中，将字段“部”直接拖放到另一个字段“组”上（字段的放置顺序会影响上下级关系，可进行拖放调整），会弹出窗口，在窗口中键入名称“组织”，单击“确定”，如图6-2所示。



图6-2　拖动方式，创建“组织”分层结构

字段“中心”也可以拖放到“组织”分层结构中，最终通过调整得到“组织”的分层结构：“中心–部–组”。

当待分层字段出现在文件夹内部时，不能再通过拖放的方式来创建分层结构，如图6-3所示。

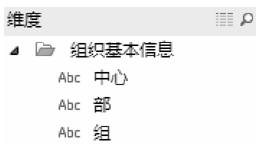


图6-3　在“组织基本信息”文件夹中不能拖动创建分层结构

2. 方式2：通过右键菜单创建名为“组织”的分层结构

在“维度窗口”中，单选或复选目标字段，右键选择“创建分层结构”，出现命名提示后，为该分层结构键入名称“组织”，单击“确定”，如图6-4所示。

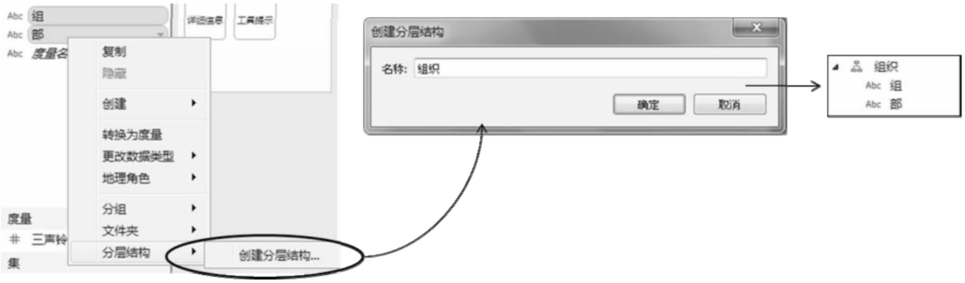


图6-4 右键菜单直接创建“组织”分层结构

创建好分层结构“组织”后，仍然可以通过拖放位置的方式调整顺序和添加新的层级。当然，还可以通过右键菜单“添加到分层结构”将遗漏的层级“中心”加入到 “组织”中，如图6-5所示。



图6-5 添加到已有分层结构

> **说明** 在Tableau中，可以通过对分层结构中的维度进行拖放和右键操作将其从分层结构中移除，也可将整个分层结构移除。当所有的层级都从分层结构中移除时，整个分层结构也就被移除了。

6.1.2 使用分层结构

在Tableau中，有两种方法可以进行上钻和下钻，一种是单击功能区字段前方的+或-，另一种是在视图标题上右键选择钻取分层结构。

1. 使用行功能区或列功能区字段进行钻取

根据上一节中创建的“组织”分层结构，由“中心”下钻到“部”的示例如图6-6所示，单

击分层结构上的“加号/减号”符号可以轻松完成钻取工作。不论在哪里使用分层结构（行功能区、列功能区或“标记”卡），一般而言遇到“加号/减号”即可进行钻取操作（加号和减号分别对应下钻和上钻）。

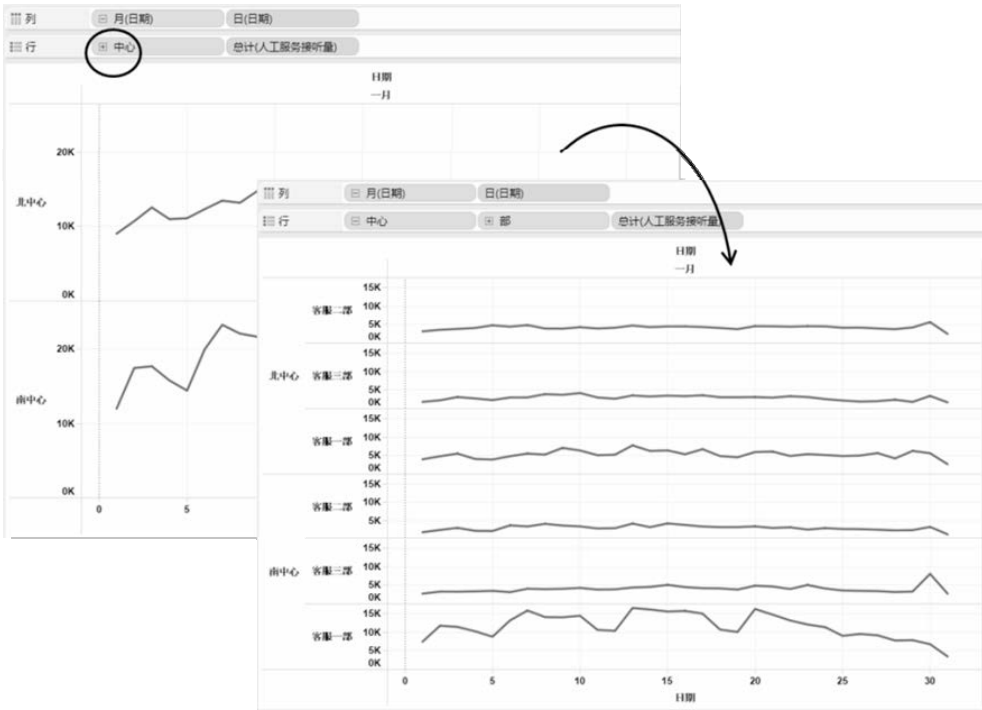


图6-6　使用行功能区上的“加号/减号”进行钻取

2. 使用视图中的标题进行钻取

另外，通过视图中的标题也可以进行上钻和下钻操作。该方法有两种方式：①右键单击视图标题，然后从上下文菜单中选择“下钻查询”或“上钻查询”（如图6-7所示）；②让鼠标在视图标题上悬浮一会儿，会显示“加号/减号”符号，便可进行上钻和下钻（如图6-8所示）。

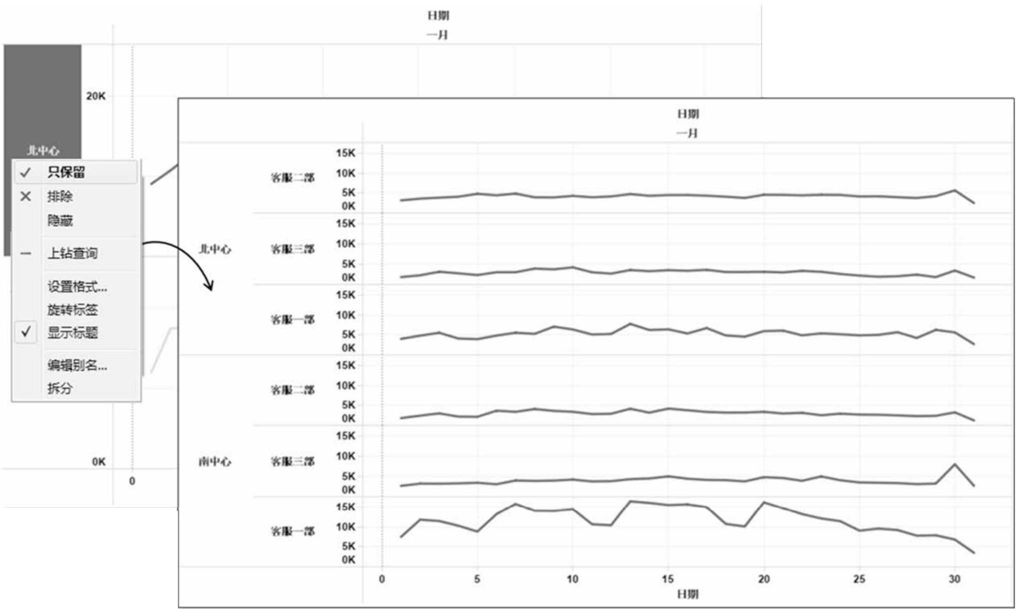


图6-7　右键单击视图中的标题“下钻/上钻查询”

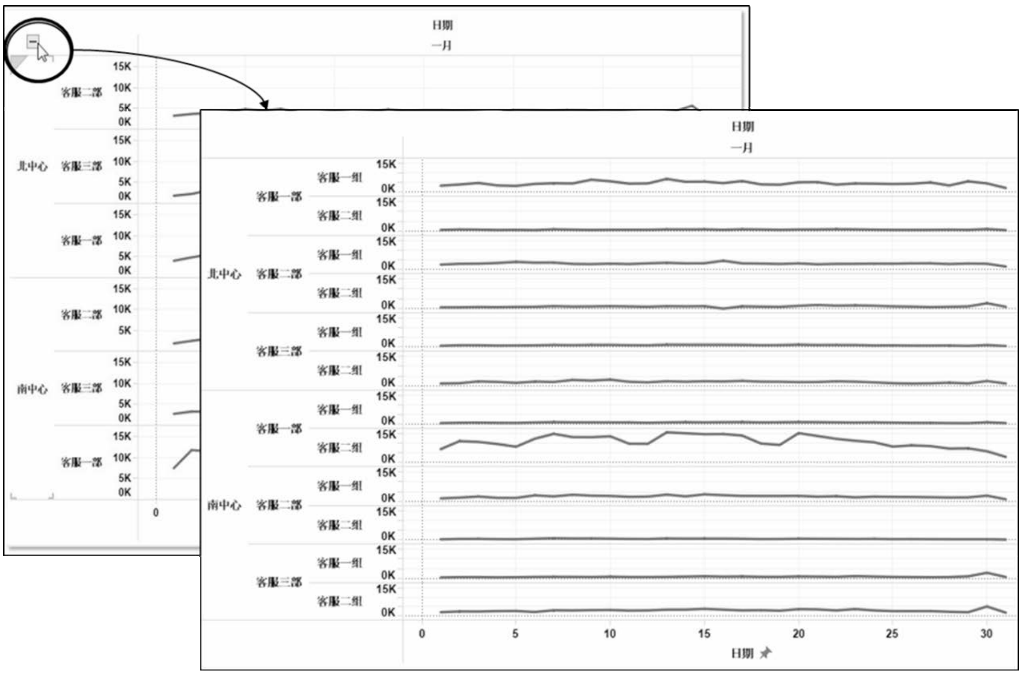


图6-8　鼠标悬停于视图标题上方，显示“加号/减号”

6.2　组

组（group）是维度成员或者度量的离散值的组合。通过分组，可以实现对维度成员的重新组合，以及度量值的按范围分类，组字段的图标为📎。

在Tableau中，要归类重组维度成员有多种方式，分组是最常见和最快速的方式之一。但是，组是不能用于计算的（计算字段会在6.5节中详细介绍），即组不能出现在公式中。

6.2.1　创建组

在坐席接听统计数据中，有些"班"名称不同，但实际为一个班，这时就可以创建组，从而对这些班进行合并处理。

组有两种创建方式：①直接在视图中选择维度成员来创建组；②基于数据窗口中某个维度来创建组。

1. 直接在视图中选择维度成员创建组

在工作表视图中，按住Ctrl单击选中维度成员，然后在选中区域悬停选择📎来创建新的组，或者在选中区域单击鼠标右键，弹出菜单，单击"组"进行创建。如图6-9所示，"班"字段包含了"13班"和"13班（15批新人）"，这两类其实都是"13班"，选择📎创建组，默认的组名称是"13班 & 13班（15批新人）"，可重命名为"13班"。

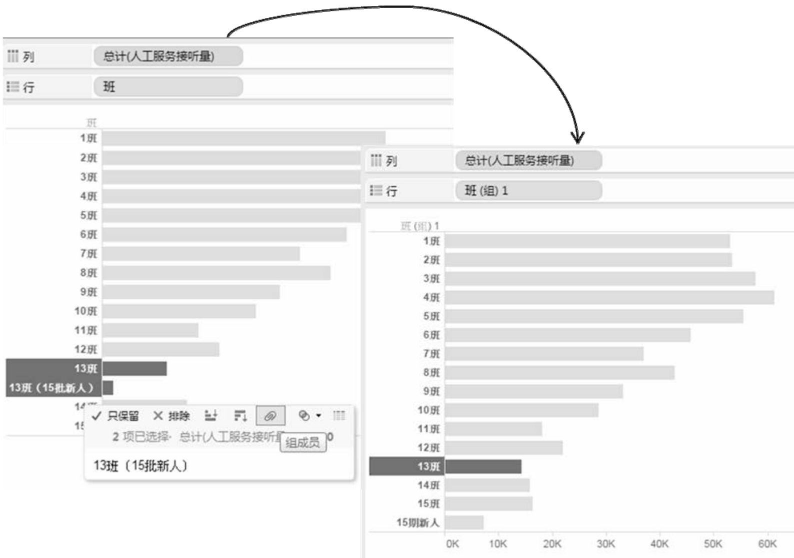


图6-9　在视图中选择维度成员创建组

当需要取消分组时，只要选择要取消分组的一个或多个维度成员，然后单击工具栏上的"取消分组"选项即可，如图6-10所示。或单击右键，在弹出对话框中单击"取消分组"。



图6-10 取消分组

2. 通过右键菜单创建分组

另一种常用的分组方式就是在维度窗口中，右键“班”字段选择“创建组”，通过这种方式可以进入“创建组[班]”界面（该界面同“编辑组[班（组）]”一致），如图6-11所示。



图6-11 右键菜单创建组

在“创建组[班]”界面中，我们发现“常白1班”“常白2班”等成员可统称为“常白班”，选中这些成员单击“分组”命名为“常白班”（还可使用“添加到”将遗漏成员添加到“常白班”中），如图6-12所示。



图6-12 创建“常白班”分组

当维度中的成员非常多时，为了更快更准确地创建分组，可以使用Tableau提供的关键字查找方法进行快速分组。如把班级名称中包含“运行”关键字的班级进行分组，可以采用如下步骤。

(1) 单击对话框底部的“查找”按钮，显示查找选项。

(2) 在“查找成员”文本框中输入要查找的成员名称的全部或一部分，这里输入“运行”。

(3) 从下拉菜单中选择适当的索引方式，有“包含”“开头为”“精确匹配”3种方式可供选择，此处选择“包含”。从“范围”中选择待查找的维度成员范围，此处选择“全部”。

(4) 单击“查找全部”按钮。

(5) 单击“分组”并重命名为“运行班”就能得到新的成员结构，如图6-13所示。

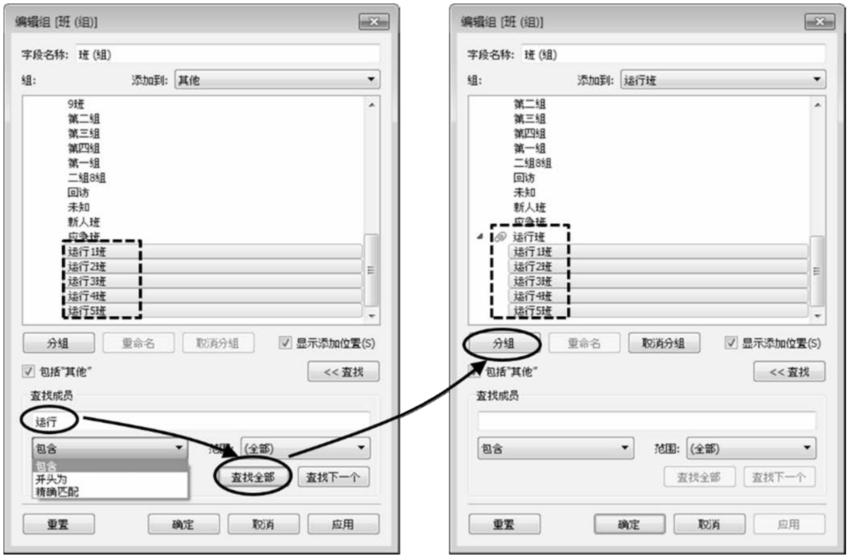


图6-13　利用关键字查找进行分组

6.2.2　使用组

在上一节中，已经对维度“班”中的成员进行了分组，得到了新的“班（组）”字段。本节将介绍“组”的两种使用方式。

1. 展示所有成员

创建后的组的“班（组）”字段包含了已分组的成员和未分组的成员，在绘制视图时，默认将未分组的成员和已分组的成员同时展示，如图6-14所示，图中既展示了“常白班”“运行班”“13班”“15班”等已分组成员，也展示了“1班”“2班”等未分组成员。

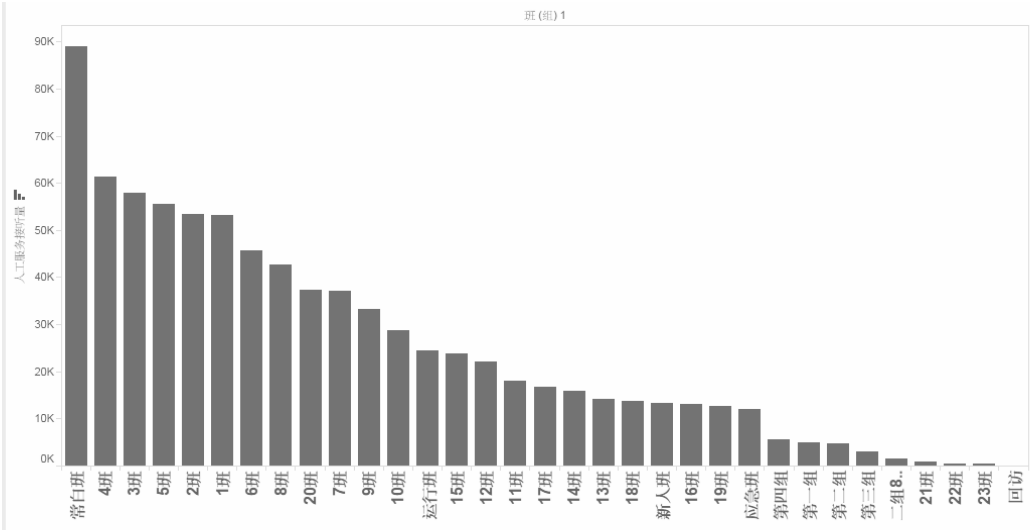


图6-14　展示“班（组）”所有成员

2. 仅展示定义好的组成员

单击列功能区中“班（组）”的下拉菜单，选择“包括其他”选项，如图6-15所示，这样“班（组）”就分为“常白班”“运行班”“15班”“13班”和“其他”，即把未定义分组的成员默认分组为“其他”。

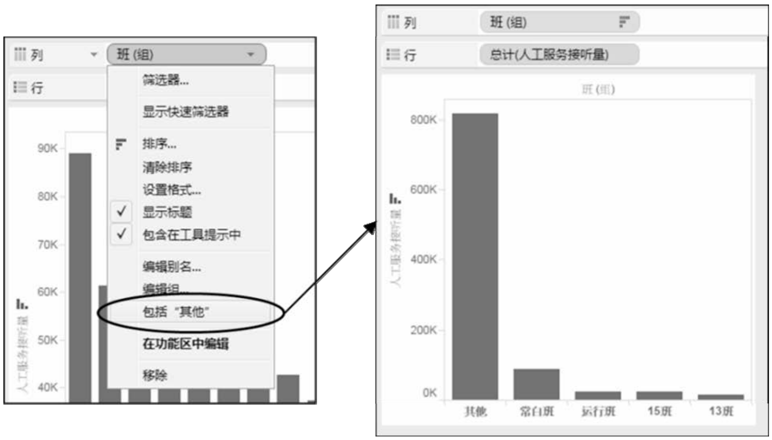


图6-15　设置仅展示定义好的组成员

Tableau也可通过在“维度窗口”或工作表中右键单击相应字段，选择“编辑组”，然后在对话框中勾选“包括其他”选项，实现仅展示定义好的组成员。

6.3　集

集（set）是根据某些条件定义数据子集的自定义字段，可以理解为维度的部分成员。Tableau在数据窗口底部显示集，并使用作为图标。集能够用于计算，参与计算字段的编辑。

1. 集的分类

根据是否能够随着数据动态变化，集可以分为两大类：常量集和计算集。其中常量集为静态集，不能跟随数据动态变化；计算集为动态集，可以跟随数据动态变化。一般情况下，集的创建针对一个维度进行，但是常量集可以是多个维度的数据子集。常量集与计算集的区别见表6-1。

表6-1　集的分类与比较

	常量集	计算集
随着数据变化	否，静态集	是，动态集
允许使用的维度数量	单个或多个维度	单个维度
创建方式	在视图中直接选择对象创建集	数据窗口右键单击维度创建集

多个集之间可进行合并操作，合并后的集为合并集。

2. 集的作用：选取维度部分成员

集主要用于筛选，通过选取维度的部分成员作为数据子集，以实现对不同对象的选取。集主要有以下两个用处。

(1) 集内外成员的对比分析。Tableau提供了集的一对特性——内/外（in/out），通过选择“在集内/外显示”可以直接对集内、集外成员进行聚合对比分析。

(2) 集内成员的对比分析。当重点为对集内成员的分析时，可选择“在集内显示成员”，此时集的作用就是筛选器，只展示位于集内的成员。

6.3.1　创建集

本节采用坐席接听统计数据，分别介绍如何创建常量集和计算集，以及创建合并集的方式。

1. 创建常量集：“平均每日人工服务接听量”由高到低排名前10名员工

首先创建基本视图，将“人工服务接听量” 拖放到“列功能区”并做平均值聚合运算，将“工号”拖放到“行”功能区，然后按照如下步骤快速创建常量集。

(1) 在视图中，按照“平均每日人工服务接听量”排序，采用降序排列，再用鼠标拖选前10名员工。

(2) 在选中区域悬停，在弹出的工具提示上，单击“创建集”选项。

(3) 在弹出的“创建集”对话框中，键入名称“平均每日人工服务接听量降序排名前10名员工”，单击“确定”，如图6-16所示。

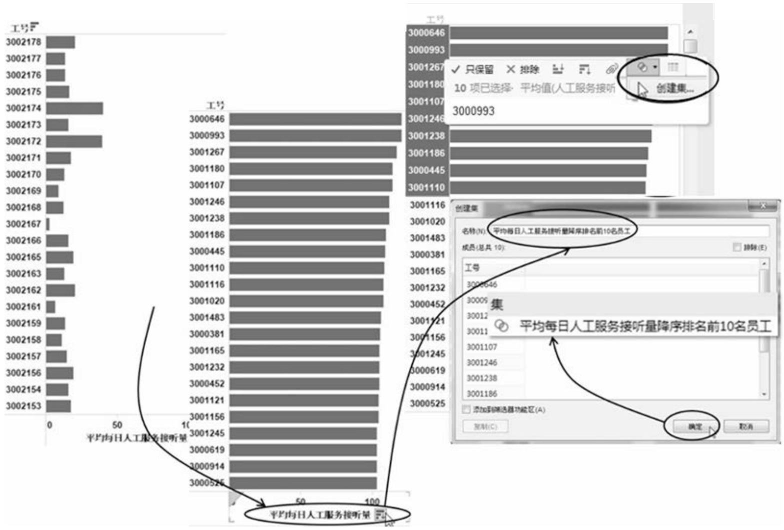


图6-16 创建“平均每日人工服务接听量降序排名前10名员工”常量集

2. 创建计算集：“出勤天数”由高到低排名前1000名员工

在创建计算集之前，首先梳理员工“出勤天数”的计算方式。数据中每一行记录是特定“工号”员工在某一天的坐席接听统计数据，那么该员工的出勤天数就是该工号在所有记录中出现的总行数。据此，“出勤天数”由高到低排名前1000名员工集的创建步骤如下。

(1) 右键“维度窗口”中的“工号”，选择“创建集”。

(2) 在弹出的“创建集”对话框中，键入名称“出勤天数降序排名前1000名员工”，并在“常规”选项卡中选择 “使用全部”。

(3) 单击“顶部”选项卡进行设置，选择“按字段”➤“顶部”➤“1000”，以及“工号”➤“计数”，然后单击“确定”，即创建了“出勤天数”由高到低排名前1000名员工集，如图6-17所示。



图6-17 创建“出勤天数降序排名前1000名员工”计算集

计算集还可按照“条件”进行设置，以实现对某个字段的值进行筛选，具体方法可参考2.3.3节筛选器的介绍。

通过以上创建过程可见，计算集对大量数据创建更为方便，同时能随着导入数据的变化动态变化，而常量集不论导入数据如何变化都是所选择的固定成员。

3. 创建合并集：高出勤且高人工服务接听量的员工

在Tableau中，集的合并要遵循相同维度，比如“平均每日人工服务接听量降序排名前10名员工”和“出勤天数降序排名前1000名员工”两个集都是以员工为维度进行筛选。维度不相同的两个集原则上不能合并，即使合并成功（有些特殊情况）也容易造成误解。

集的合并有3种方式：①并集，包含两个集内的所有成员；②交集，仅包含两个集内均存在的成员；③差集：包含指定集内存在而第二个集内不存在的成员，即排除共享成员。

当上述两个集准备完毕后，按照如下步骤创建名为 “高出勤且高人工服务接听量的员工”的合并集，即出勤天数前1000名且平均每日人工服务接听量为前10名的员工。

(1) 在“数据窗口”中选择要合并的两个集“平均每日人工服务接听量降序排名前10名员工”与“出勤天数降序排名前1000名员工”，右键菜单选择“创建合并集”，如图6-18所示。



图6-18　创建合并集“高出勤且高人工服务接听量的员工”

(2) 在“创建集”对话框中，键入新创建的合并集的名称：“高出勤且高人工服务接听量的员工”，确认要合并的两个集在两个下拉菜单中都处于选中状态，然后选择合并方式，此处选择“两个集中的共享成员”，最后单击“确定”即可创建合并集。

6.3.2　使用集

1. 集内外成员对比分析

下面以分析“高出勤且高人工服务接听量的员工”（以下简称“勤劳员工”）在各个中心、组的分布情况为例，介绍集内外成员的对比分析方法。

首先创建一个各中心、组的员工人数柱图，将“中心”“组”拖到列功能区，“工号”按照“计数（不同）”作聚合后拖到行功能区。然后将集“出勤天数降序排名前1000名”拖放到“颜色”。此时，各组内的“勤劳员工”和“其他”员工以不同颜色对比展示，编辑颜色图例别名，集内成员为“勤劳员工”，集外成员为“其他”。将“工号”作聚合后拖到“标签”，选择快速表计算①为“总额百分比”，并编辑表计算依据为“高出勤且高人工服务接听量的员工 内/外”。此时柱图上的百分比标签为各组内“勤劳员工”所占比例，可看出“勤劳员工”全部分布在南中心的客服二组，占客服二组全部员工的1.83%，如图6-19所示。



图6-19　高出勤且高人工服务接听量的员工占比分布

2. 各组内“勤劳员工”占比对比分析集内成员对比分析

下面以分析南北中心在“出勤天数降序排名前1000名员工” 中的比例为例，介绍集作为筛选器的使用方法。

首先创建一个南北中心员工人数占比的饼图，把“中心”拖放到“颜色”和“标签”，把“工号”拖放到“角度”和“标签”处并按“计数（不同）”聚合。然后将集“出勤天数降序排名前1000名”拖放到筛选器，即显示 “出勤天数降序排名前1000名员工”在南北中心的占比的饼图，如图6-20所示。

① 快速表计算将在6.5.4节进行详细介绍。



图6-20　南北中心“出勤天数降序排名前1000名员工”人数占比分析

6.4　参数

参数（parameter）是一种可用于交互的动态值。Tableau 在数据窗口底部显示参数，并使用图标 # 作为标签。

参数是由用户自定义的动态值，是实现控制与交互的最常见、最方便的方法，被广泛地运用在可动态交互的字段（计算集、自定义计算字段等）、筛选器及参考线（包括参考区间等），分析人员可以轻松地通过控制参数来与工作表视图进行交互。如图6-21所示，通过参数控件，可以调整其他字段，进而控制工作表视图。参数在工作簿中是全局对象，可在任何工作表中单独使用，也可同时应用于多个工作表视图。

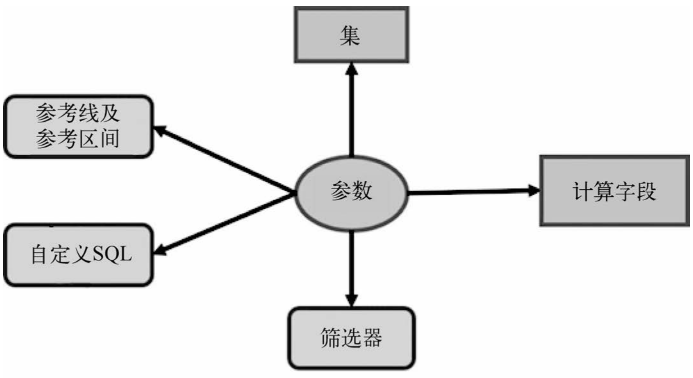


图6-21　参数的使用

本节主要介绍参数与计算集的配合使用，参数与计算字段的配合使用案例将在6.5.3节中详细介绍，参数与参考线的共同使用案例将在6.7.2节中展开介绍。

6.4.1 创建参数

参数的创建方式有多种，但总体来说可以归纳为两类：①直接在数据窗口中创建；②在使用计算集、计算字段、参考线及其他功能时创建。

1. 直接在数据窗口创建参数："服务评价满意率阈值"

在数据窗口中创建参数的步骤如图6-22所示。



图6-22 数据窗口创建"服务评价满意率阈值"参数

在数据窗口中，右键单击作为参数基础的字段"服务评价满意率"，弹出菜单选择"创建"➤"参数..."，也可以在"数据窗口"右上角的下拉箭头中打开菜单选择"创建参数"。

在弹出的对话框中，设置参数的名称、注释和属性。

(1) 名称：输入想设置的参数名称"服务评价满意率阈值"。

(2) 注释：输入对参数意义的描述，以帮助其他人理解所设参数的含义，此处非强制项，可不设置。

(3) 属性："数据类型"用于指定参数将接受的值的数据类型。"当前值"用于指定参数的默认值。"显示格式"选项用于指定要在参数控件中数值的显示格式。"允许的值"选项用于指定参数接受值的方式，包括3种类型，①"全部"表示参数可调整为任意值；②"列表"表示参数设置为列表内的值，有3种设置方法，分别手动输入、从字段中添加或从剪切板粘贴；③"范围"表示参数可在指定值范围内进行调整，可设置最小值、最大值和每次调整的步长，也可从参数设置或从字段设置。一般情况下，作为参数基础的字段是维度时，"允许的值"表现为列表，当作为参数基础的字段是度量时，"允许的值"表现为范围。

Tableau默认的参数属性“数据类型”为“浮点”，“当前值”为0，“允许的值”是在0~1之间的“范围”。如果不符合实际需求，可随意调整，本处把步长设置为0.1，把“显示格式”由“自动”调整为“百分比”，并设置展示两位小数。单击 “确定”，“服务评价满意率阈值”参数就会显示在数据窗口中。

2. 在使用计算集时创建参数

在计算集中创建参数的步骤如图6-23所示。

(1) 右键单击“出勤天数降序排名前1000名员工”计算集，在“编辑集”窗口中，修改集名字为“出勤天数降序排名前N名员工”，在输入数值的下拉菜单中，选择 “创建参数”。

(2) 在弹出的对话框中，设置参数的名称、注释、属性，与直接在“数据窗口”中创建参数的方法一致。参数名称为“出勤天数降序TopN员工阈值”，数据类型设置为“整数”，“允许的值”为“范围”，设置为1~3000，步长为1，单击“确定”即成功创建参数。

注意，当“步长”处于非激活状态时，Tableau会根据数据范围自动选择相应的步长。



图6-23　编辑计算集时创建“出勤天数降序TopN员工阈值”参数

6.4.2　使用参数

在6.3.2节中对“出勤天数降序排名前1000名的员工”进行了南北中心的对比分析，如需动态查看出勤天数排名不同的员工数量对比，需要引入参数进行手动改变，设置步骤如下。

(1) 在数据窗口中右击参数“出勤天数降序TopN员工阈值”，并选择“显示参数控件”，此时参数控件将显示在视图区域的右上角。

(2) 单击参数控件的下拉箭头可设置参数的展示形式，包括“编辑标题”“设置参数格式”“滑块”“键入内容”等，其中“设置参数格式”可调整参数标题、正文的字体格式和大小等；当选择“滑块”时，可通过“自定义”选择是否“显示读出内容”“显示滑块”和“显示按钮”；当选择“键入内容”时，只展示读出内容。“滑块”和“键入内容”的展示形式如图6-24所示。



图6-24 设置参数控件显示形式

(3) 将集“出勤天数降序排名前N名员工”拖入筛选器，调整参数的值，可动态观察不同排名的员工数量在南北中心的分布，如图6-25所示。

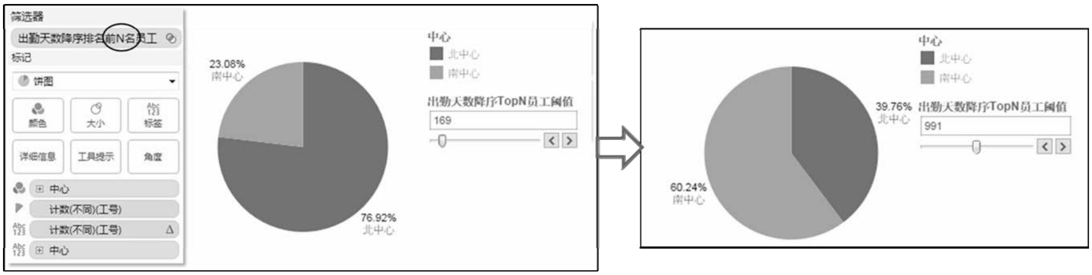


图6-25 使用参数控件动态调整视图

6.5 计算字段

计算字段（calculated field）是根据数据源字段（包括维度、度量、参数等）使用函数和运算符构造公式来定义的字段。同其他字段一样，计算字段也能拖放到各功能区来构建视图，还能用于创建新的计算字段，而且其返回值也有数值型、字符型等的区分。

计算字段的创建界面如图6-26所示，包括了输入窗口和函数窗口。

在输入窗口中，可输入计算公式，包括运算符、计算字段和函数。其中，运算符支持加（+）、减（–）、乘（*）、除（/）等所有标准运算符。字符、数字、日期/时间、集、参数等字段均可作为计算字段，Tableau的自动填写功能会自动提示可使用计算的字段或函数，如图6-26所示。例如输入“不”之后，Tableau会提示有度量“不满意评价数”和集“不满意评价数低于5的员工”供

选择，也可选择orgin数据源中的度量“不满意数”。

函数窗口为Tableau自带的计算函数列表，包括数字、字符串、日期、类型转换、逻辑、聚合以及表计算等，详细介绍参见附录。双击该函数即可在“输入窗口”中出现，也可在“输入窗口”中自动补全。



图6-26　计算字段创建界面

6.5.1　创建计算字段

同参数的创建类似，计算字段的创建方式有两种：①直接在数据窗口中创建计算字段；②在使用计算集、计算字段、参考线及其他时创建。本节主要介绍如何通过“数据窗口”创建计算字段的方法。

1. 创建一个简单的“服务评价满意数”计算字段

按照业务逻辑，“服务评价满意数”为 “服务评价推送成功数”和“不满意评价数”的差。创建计算字段只需在数据窗口中的字段上单击右上角小箭头，或右键单击字段，在弹出菜单中选择“创建”➤“计算字段”，如图6-27所示，在弹出的对话框中输入公式“[服务评价推送成功数]-[不满意评价数]”，单击“应用”，这样Tableau就将对每一条记录按照该公式进行计算，并生成一个新的字段“服务评价满意数”。



图6-27　创建计算字段“服务评价满意数”

因为该计算返回数字，所以新字段显示在数据窗口的“度量”区域中，并且可以像使用其他任何字段那样使用该新字段。

在公式输入框中，可以使用“//”开头来书写注释。

2. 运用逻辑函数与参数创建“服务评价满意率分类”计算字段

使用参数和Tableau的逻辑函数创建计算字段，生成“满意”和“不满意”两个类别，实现对服务评价满意率的分级，通过调节参数“服务评价满意率阈值”，当“服务评价满意率”大于参数阈值时，则服务分级为“满意”，否则为“不满意”，实现灵活分级。

采用Tableau的逻辑函数IF语句，具体函数使用方式可参考窗口右方的函数描述和示例，计算字段语句为：“`IF 服务评价满意浪费>服务评价满意率阈值 THEN '满意' ELSE '不满意' END`”，具体创建方式如图6-28所示。公式输入后，在左下角会提示是否正确，当显示“计算有效”时，单击“确定”，即生成“服务评价满意率分类”字段。



图6-28　创建计算字段“服务评价满率分类”

根据返回的数据类型，该计算字段自动显示在数据窗口的“维度”中，使用=Abc作为标签，如图6-29所示。

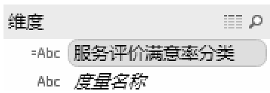


图6-29　“服务评价满意率分类”字段

6.5.2　使用计算字段

使用计算字段“服务评价满意率分类”与参数“服务评价满意率阈值”来创建服务评价满意率分析视图，实现对每个员工每天的话务接听量和服务满意率进行综合评价，创建步骤如下。

(1) 将“服务评价满意率”以及“人工服务接听量”分别拖放到行、列功能区，并选择标记类型为“圆”。由于默认是聚合状态，而此处要分析的是每人每天的业务情况，即每条记录的分布，所以取消菜单栏“分析”中的“聚合度量”选项，进行解聚。

(2) 将计算字段“服务评价满意率分类”拖放到“颜色”，并且显示参数控件“服务评价满意率阈值”。如图6-30所示，可以调整阈值来观察不同阈值条件下每个员工每天的“人工服务接听量”与“服务评价满意率”之间的关系和变化。

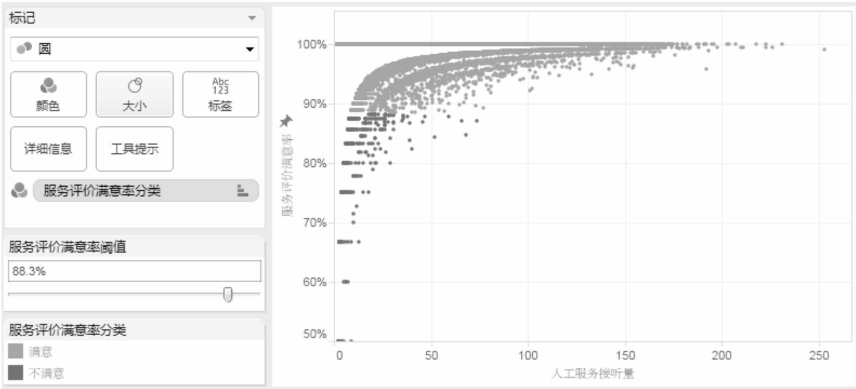


图6-30　服务评价满意率分类

为了能够更快速地看到分析结果，Tableau提供了在行功能区和列功能区直接输入计算公式的方式，这种方式下创建的计算字段可即时在视图中看到结果，拖放该字段到数据窗口，即可形成新的字段。

如将图6-30中的人工服务接听量的计量单位由“个”修改为“百个”，则在列功能区中双击“人工服务接听量”，直接输入“/100”，此时视图中的X轴即修改为“SUM([人工服务接听量])/100”，如图6-31所示。

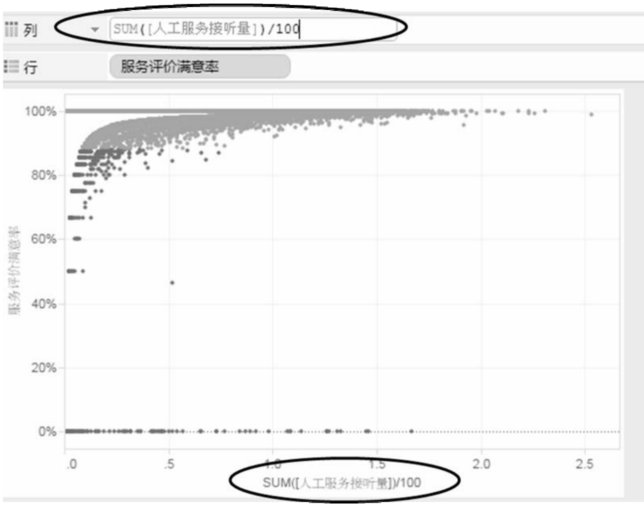


图6-31　服务评价满意率分类

6.5.3 特殊函数：表计算

表计算是针对数据库中多行数据进行计算的方式。当创建表计算后，在标记卡、行功能和列功能区域，该计算字段就会有正三角标记，如 总计(人工服务接听量) Δ 。

表计算函数针对度量使用 “分区”和“寻址”进行计算，这些计算依赖于表结构本身。在编辑公式时，表计算函数需要明确计算对象和使用的计算类型。而最需要注意的是，在使用表计算时必须使用聚合数据。

1. 快速表计算

Tableau把常用的表计算嵌入到“快速表计算”中，利用它们能非常快速地使用表计算结果。

如图6-32所示，右键单击“总计（人工服务接听量）”或选择下拉箭头，在弹出的菜单中选择“快速表计算”➤“差异”。或直接单击 “添加表计算”，在弹出的对话框中选择“差异”。此时，默认表计算的逻辑是沿着“表（横穿）”相对于“上一个”顺次计算差值，如4840－4029得到811，5561－4840得到721，以此类推。双击该字段，可看到快速表计算的公式为`ZN(SUM([人工服务接听量])) – LOOKUP(ZN(SUM([人工服务接听量])), -1)`。



图6-32 快速表计算（差异）（按时间递延计算相邻接听量的差异）

如果希望获得与“第一个”值（即与该月1日数据）的差异，单击“编辑表计算”在“将值显示为以下项的差异”下拉列表中选择“第一个”，如图6-33所示。或直接单击菜单中的“相对于”，将其修改为“第一个”即可。双击该字段，可看到快速表计算的公式为`ZN(SUM([人工服务接`

听量])) - LOOKUP(ZN(SUM([人工服务接听量])), FIRST())。



图6-33　编辑表计算

这样就可以获得每一个值与第一个值的差值，如图6-34所示。



图6-34　编辑表计算（沿着日期对每个“中心–部”下人工服务接听量计算其与第一个的差异）

在高级分析中，“快速表计算”是比较常用的方式。Tableau共嵌入了包括汇总、差异、百分比差异、总额百分比、排序、百分位、移动平均、YTD总计（本年迄今总计）、复合增长率、年同比增长和YTD增长（本年迄今增长）共计11个快速表计算，可实现对表中一组数据的快速计算总计、差异、移动平均等。

除了快速表计算，Tableau还提供了多种表计算函数，可以使用它们来灵活编辑公式，自定义表计算。在介绍自定义表计算之前，首先介绍两个重要概念：寻址字段和分区字段。

2. 寻址和分区

在快速表计算中，可以设置计算范围，来定义表计算的计算依据，如图6-35所示。



图6-35　设置差异计算范围

单击下拉菜单中“高级”，在弹出的对话框中，对“寻址”和“分区”进行设置可自定义表计算的计算依据，如图6-36所示。



图6-36　高级设置（寻址–分区–排序）

寻址指的是对计算对象进行定义的维度字段，确定用于表计算的对象，可以按照不同深度级别进行寻址，通过“上移”或“下移”调整寻址的先后顺序。一般来说，寻址可以相对于表结构（以“表”“区”或“单元格”开头的选项）或相对于特定字段（例如中心、部、组或日期等）。分区则是对计算对象进行分组的维度字段，确定计算时的分组方式。排序是确定各个分区的计算对象的计算顺序。

3. 计算依据设置

计算依据设置主要有以下3个。

● **表（横穿）及表（向下）**

表（横穿）可以理解为对每一个分区沿着水平方向进行特定的计算，即将寻址设置为计算整个表，并且沿水平方向移动计算每个分区，这是Tableau默认的计算依据。

如图6-37所示，表上横向排列的维度是寻址字段，即“日期”，而所有其他维度（“中心”“部”）均为分区字段。

		日期									
中心	部	1	2	3	4	5	6	7	8	9	10
北中心	客服一部	4,029	4,840	5,561	4,128	3,965	4,863	5,563	5,290	7,096	6,389
	客服二部	3,183	3,611	3,849	4,095	4,813	4,443	4,870	3,950	3,922	4,302
	客服三部	1,833	2,267	3,147	2,763	2,330	3,025	3,032	3,940	3,748	4,235
南中心	客服一部	7,461	11,773	11,452	10,262	8,804	13,195	15,855	14,078	14,021	14,441
	客服二部	1,829	2,445	3,016	2,215	2,159	3,688	3,404	4,105	3,642	3,414
	客服三部	2,720	3,272	3,245	3,330	3,480	3,118	4,047	3,936	4,027	4,270

图6-37　表（横穿）

同理，表（向下）可以理解为对每一个分区沿着垂直方向进行特定的计算，如图6-38所示。此时其寻址字段为“中心”“部”，分区字段为“日期”。

		日期									
中心	部	1	2	3	4	5	6	7	8	9	10
北中心	客服一部	4,029	4,840	5,561	4,128	3,965	4,863	5,563	5,290	7,096	6,389
	客服二部	3,183	3,611	3,849	4,095	4,813	4,443	4,870	3,950	3,922	4,302
	客服三部	1,833	2,267	3,147	2,763	2,330	3,025	3,032	3,940	3,748	4,235
南中心	客服一部	7,461	11,773	11,452	10,262	8,804	13,195	15,855	14,078	14,021	14,441
	客服二部	1,829	2,445	3,016	2,215	2,159	3,688	3,404	4,105	3,642	3,414
	客服三部	2,720	3,272	3,245	3,330	3,480	3,118	4,047	3,936	4,027	4,270

图6-38　表（向下）

● **表（横穿，然后向下）及表（向下，然后横穿）**

表（横穿，然后向下）将寻址设置为先横向后竖向计算整个表，即表中横向和竖向排列的字段都是寻址字段，其顺序为“中心”➤“部”➤“日期”，但没有分区字段，其计算过程为：横向计算，移至下一行继续横向计算，以此类推，如图6-39所示。

		日期									
中心	部	1	2	3	4	5	6	7	8	9	10
北中心	客服一部	4,029	4,840	5,561	4,128	3,965	4,863	5,563	5,290	7,096	6,389
	客服二部	3,183	3,611	3,849	4,095	4,813	4,443	4,870	3,950	3,922	4,302
	客服三部	1,833	2,267	3,147	2,763	2,330	3,025	3,032	3,940	3,748	4,235
南中心	客服一部	7,461	11,773	11,452	10,262	8,804	13,195	15,855	14,078	14,021	14,441
	客服二部	1,829	2,445	3,016	2,215	2,159	3,688	3,404	4,105	3,642	3,414
	客服三部	2,720	3,272	3,245	3,330	3,480	3,118	4,047	3,936	4,027	4,270

图6-39　表（横穿，然后向下）

同理，表（向下，然后横穿）的寻址字段依次为“日期”➤“中心”➤“部”，没有分区字

段，两者的区别在于寻址字段的顺序，如图6-40所示。

中心	部	日期 1	2	3	4	5	6	7	8	9	10
北中心	客服一部	4,029	4,840	5,561	4,128	3,965	4,863	5,563	5,290	7,096	6,389
	客服二部	3,183	3,611	3,849	4,095	4,813	4,443	4,870	3,950	3,922	4,302
	客服三部	1,833	2,267	3,147	2,763	2,330	3,025	3,032	3,940	3,748	4,235
南中心	客服一部	7,461	11,773	11,452	10,262	8,804	13,195	15,855	14,078	14,021	14,441
	客服二部	1,829	2,445	3,016	2,215	2,159	3,688	3,404	4,105	3,642	3,414
	客服三部	2,720	3,272	3,245	3,330	3,480	3,118	4,047	3,936	4,027	4,270

图6-40 表（向下，然后横穿）

- **区（向下）**

区（向下）将对表中的区向下进行计算，其中，“中心”“日期”为分区字段，“部”是寻址字段，如图6-41所示。

中心	部	日期 1	2	3	4	5	6	7	8	9	10
北中心	客服一部	4,029	4,840	5,561	4,128	3,965	4,863	5,563	5,290	7,096	6,389
	客服二部	3,183	3,611	3,849	4,095	4,813	4,443	4,870	3,950	3,922	4,302
	客服三部	1,833	2,267	3,147	2,763	2,330	3,025	3,032	3,940	3,748	4,235
南中心	客服一部	7,461	11,773	11,452	10,262	8,804	13,195	15,855	14,078	14,021	14,441
	客服二部	1,829	2,445	3,016	2,215	2,159	3,688	3,404	4,105	3,642	3,414
	客服三部	2,720	3,272	3,245	3,330	3,480	3,118	4,047	3,936	4,027	4,270

图6-41 区（向下）

以差异计算为例，计算依据为区（向下）相当于对每一个分区单独进行向下的差异计算，其结果如图6-42所示。

中心	部	日期 1	2	3	4	5	6	7	8	9	10
北中心	客服一部	4,029	4,840	5,561	4,128	3,965	4,863	5,563	5,290	7,096	6,389
	客服二部	3,183	3,611	3,849	4,095	4,813	4,443	4,870	3,950	3,922	4,302
	客服三部	1,833	2,267	3,147	2,763	2,330	3,025	3,032	3,940	3,748	4,235
南中心	客服一部	7,461	11,773	11,452	10,262	8,804	13,195	15,855	14,078	14,021	14,441
	客服二部	1,829	2,445	3,016	2,215	2,159	3,688	3,404	4,105	3,642	3,414
	客服三部	2,720	3,272	3,245	3,330	3,480	3,118	4,047	3,936	4,027	4,270

中心	部	日期 1	2	3	4	5	6	7	8	9	10
北中心	客服一部										
	客服二部	-846	-1,229	-1,712	-33	848	-420	-693	-1,340	-3,174	-2,087
	客服三部	-1,350	-1,344	-702	-1,332	-2,483	-1,418	-1,838	-10	-174	-67
南中心	客服一部										
	客服二部	-5,632	-9,328	-8,436	-8,047	-6,645	-9,507	-12,451	-9,973	-10,379	-11,027
	客服三部	891	827	229	1,115	1,321	-570	643	-169	385	856

图6-42 区（向下）计算结果

其他和区相关的计算范围，比如区（横穿）、区（横穿，然后向下）及区（向下，然后横穿），都是针对每个区进行的计算，其区别只是寻址方式不同。

除了按照表、区进行计算依据设置外，在Tableau中还可以根据单元格、单个字段甚至多个字段进行计算。

- 单元格：当设置为“单元格”时，所有字段都是分区字段。在计算总额百分比时，此选项通常最有用。
- 单个字段：将寻址设置为指定的字段进行计算。此选项的好处是可以绝对控制计算方式，即使更改视图方向，表计算也将继续使用相同的寻址和分区字段。但请注意，对单个字段寻址意味着如果重新排列表，计算可能不再与表结构匹配。
- 多个字段（高级）：通过高级选项可以指定多个字段充当寻址字段和分区字段。如果选择“高级”则会打开一个对话框，在其中可以指定一个或多个字段充当寻址字段和分区字段，然后还可以确定这些字段的顺序。通过寻址、分区、排序3个设置，灵活地使用表计算。

4. 自定义表计算

自定义表计算就是在创建计算字段时使用表计算函数进行编辑，Tableau提供了32个表计算函数，详细见附录C.7。本节以使用WINDOW_AVG函数计算移动平均值为例，介绍自定义表计算的方法。

(1) 创建参数“N日移动平均”，数据类型为“整数”，允许的值为“范围”，最小值为1，最大值为31，步长为1，如图6-43所示。



图6-43　创建参数

(2) 创建计算字段“移动平均”，公式为WINDOW_AVG(SUM([人工服务接听量]),-[N日移动平均],0)，其中WINDOW_AVG将返回前面N行到当前行的平均值，如图6-44所示。

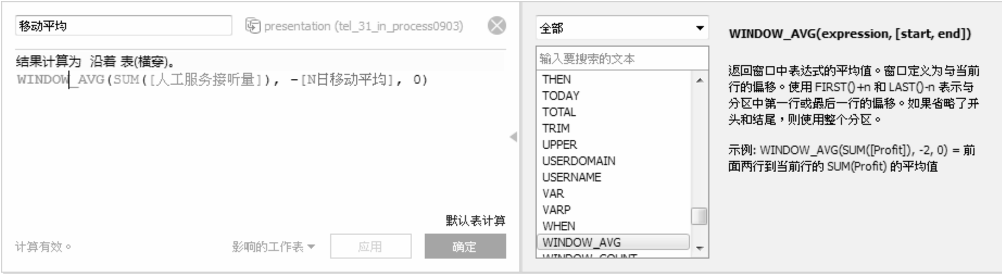


图6-44 含有参数的移动平均公式

(3) 单击“默认表计算”，进行设置，此处与快速表计算的编辑表计算功能一致。



图6-45 表计算设置

(4) 拖放“日期”和“人工服务接听量”到视图区域，生成每日人工服务接听量趋势。拖放“移动平均”到行功能区，设置计算依据为“日期”，生成双线图，如图6-46所示。

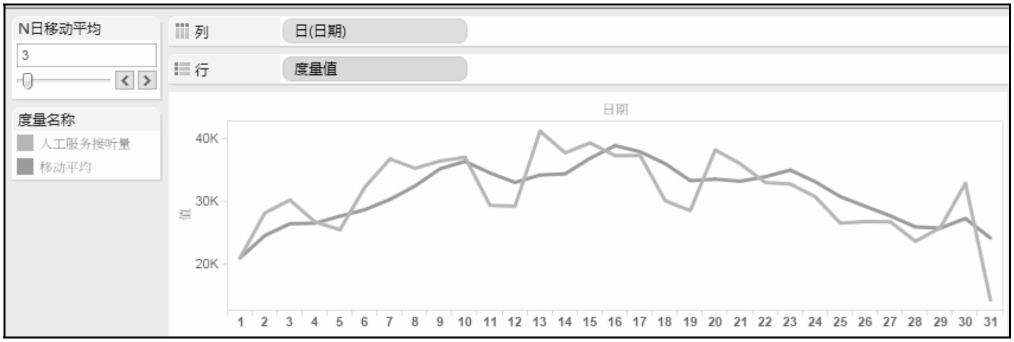


图6-46 二十日移动平均

这样就创建了以“N日移动平均”为参数进行移动平均的线图，可以通过参数动态调整移动平均的步长。

6.5.4　特殊函数：详细级别表达式

在Tableau中，我们可以通过将数据拖到视图的部分区域来实现不同明细程度的聚合与可视化展示。这些视图区域包括行功能区、列功能区，以及标记卡中的颜色、大小、标签、详细信息及路径。如果分析过程中需要添加某一维度，其明细程度高于或低于已有视图的可视化明细程度，但又不希望改变现有图形展示内容，可采用Tableau 9.0及以上版本提供的详细级别表达式功能。通过它无需将这些维度拖入已有视图中，即可实现独立于可视化详细级别，自定义数据以何种详细级别来进行计算。

详细级别表达式共有3种函数，分别是INCLUDE、EXCLUDE、FIXED，每种函数可实现不同明细程度的聚合。其中，INCLUDE函数可用于创建明细程度高于可视化展示内容的计算字段，EXCLUDE函数可用于创建明细程度低于可视化展示内容的计算字段，FIXED函数的应用不受可视化明细程度的限制，可用于创建指定明细程度的计算字段，其计算结果可以比可视化展示内容明细程度更高或更低，如图6-47所示。

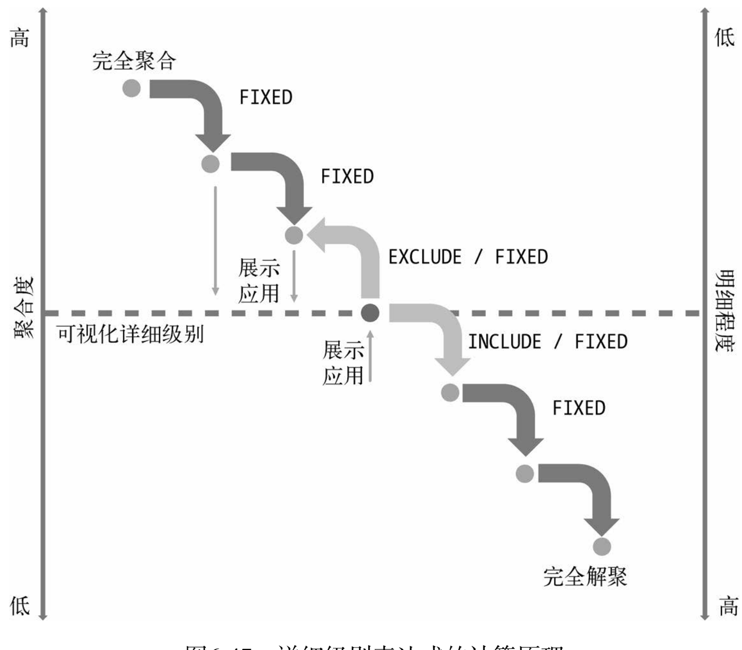


图6-47　详细级别表达式的计算原理

Tableau支持针对最新版本的Oracle、Teradata、Microsoft SQL Server、SAP HANA、Spark、MySQL等数据源创建和应用详细级别表达式，而对于Microsoft Excel、Microsoft Access和文本等数据源则不支持，如有分析需求，可先进行数据提取再创建。

另外，INCLUDE和EXCLUDE创建形成的计算字段只能当作度量使用，而FIXED函数创建形成的计算字段可用作维度，也可用作度量。

1. INCLUDE函数

以“坐席接听统计数据”为例，如果我们想了解哪个中心（地区）平均的员工服务接听量最大，需要计算出每个员工的服务接听量后再按所属地区计算平均值。在Tableau 9.0及以上版本中，借助详细级别表达式，可以轻松解决这个问题。在如下视图中，仅有[中心]一个维度，通过创建计算字段[员工服务接听量] = {INCLUDE [员工 ID]:SUM[人工服务接听量]}，可以将[员工服务接听量]纳入视图中，计算各中心（地区）每位员工的服务接听量，而INCLUDE表达式再以平均值被聚合，得到的就是该中心（地区）员工服务接听量的平均值，见图6-48。

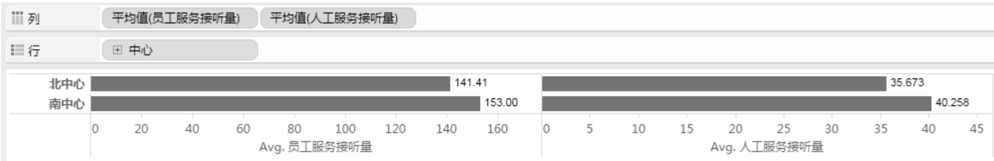


图6-48　利用INCLUDE函数计算不同地区平均的员工服务接听量

从图6-48可以看出，左侧条形图表示各中心（地区）平均的员工服务接听量（通过详细级别表达式计算得出），而右侧条形图表示各地区的平均接听量（按照数据源的最明细行项目，即接听记录进行平均值）。

2. EXCLUDE函数

以“坐席接听统计数据”为例，如果我们想了解每天的总接听量和每个中心（地区）的总接听量。为此，需要执行如下算法：在计算总接听量中将中心（地区）排除，在计算各中心（地区）接听量时将地区包括在内。在图6-49中，有[中心]和[日期]两个维度，通过添加计算字段[人工服务接听量总计] = {EXCLUDE [中心]:SUM([人工服务接听量])}可以把视图中已有的[中心]维度排除在聚合计算之外，计算的是各中心（地区）接听量的总和。

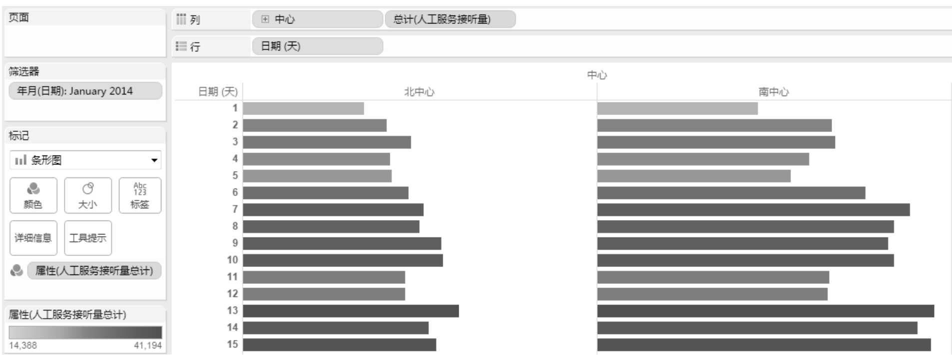


图6-49　利用EXCLUDE函数计算每日的总接听量

从图6-49可以看出，每天的总接听量（以条形图的颜色表示）依据EXCLUDE函数计算得出，按中心（地区）计算的总接听量（以条形图的长度表示）为源数据的简单聚合，于是实现了在同一可视化视图中展示了两种不同的明细程度。

3. FIXED函数

以“坐席接听统计数据”为例，创建一个名为“每位员工服务接听量”的计算字段，见图6-50。



图6-50　利用FIXED函数计算每位员工服务接听量

该表达式可指示Tableau为每个员工工号执行聚合操作，借助该表达式，可以计算每位员工的服务接听量之和。将该新字段拖入视图中后，我们就能计算出每位员工的服务接听量。

4. 详细级别表达式与表计算的区别

详细级别表达式与表计算有着不同的运算方式。 表计算完全由查询结果生成，而详细级别表达式通常是作为针对基础数据源查询的一部分而生成；表计算总是以生成度量作为结果，而详细级别表达式则能创建度量、聚合度量或维度。在Tableau 8或者更早版本的软件中，有时会用到表计算来指定计算的聚合级别。但是，由于表计算从查询结果聚合而来，所以只能生成聚合度/粒度等于或高于/低于可视化详细级别的结果。在Tableau 9.0及以上的版本中，借助详细级别表达式可以很好的解决这一问题，使得数据分析的过程更加清晰、简洁。

6.5.5　特殊函数：百分比

在使用Tableau时，我们常需要计算数据的百分比。例如，在“坐席接听统计数据”中，为了评价每位员工的工作情况，可能不需要查看每位员工的服务接听量，而只需查看每位员工的服务接听量占所有员工总接听量的百分比即可。

1. 计算百分比

进行百分比计算时，需要指定分母的数据范围。Tableau默认分母的数据范围是整个表。不过，从“分析”➤“百分比”菜单中可以自定义设置，可以修改为一行、一列或一个区等，见图6-51。

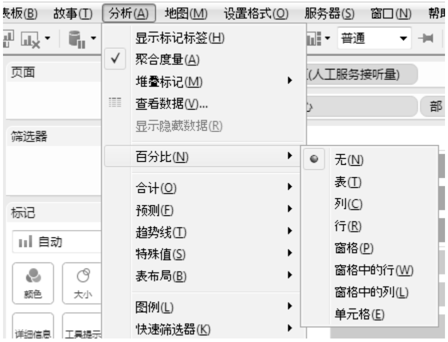


图6-51 “百分比”菜单项

2. 百分比与聚合

一般情况下，百分比是基于度量的聚合值进行计算的，在实际应用中往往容易忽略这一点。标准聚合包括总和、平均值以及若干其他聚合。例如，我们想了解每个员工“服务接听量”在总体接听量中的占比，则计算方式为该员工接听量之和（{FIXDE [员工ID]:SUM([人工服务接听量])}）除以总体接听量（SUM(人工服务接听量)）。

6.6 变换

Tableau只针对字符型字段和日期型字段具备变换功能。使用变换功能可以将单列中的复合数据拆分成多个字段，也可以将日期字段自定义为年、月、周等精度，即无需编辑数据源即可实现对数据的重新分配。

6.6.1 变换日期型字段

Tableau可针对日期进行变换，生成不同精度的日期字段或数字字段，本节以坐席接听统计数据中的“日期”为例，介绍变换的方式。

右键单击维度窗口中的“日期”字段，选择“变化”➤“创建自定义日期”，在弹出的窗口中，设置生成字段的名称、详细信息和格式，其中可设置的日期格式包括日期部分和日期值两种，如图6-52所示，选择“日期值”生成的字段仍为日期格式；选择“日期部分”生成的字段为数字格式。



图6-52 自定义拆分设置

以选择详细信息为“天”，格式为“日期部分”为例，生成的字段信息如图6-53所示，生成数型字段“日期（天）”。



图6-53　生成“日期（天）”字段

6.6.2　变换字符型字段

Tableau对字符型字段的变换包括拆分和自定义拆分两种形式，其中拆分为Tableau默认的拆分方式，可以将数字与文本进行拆分；自定义拆分可以根据数据的格式，自定义拆分方式，包括拆分使用的分隔符、拆分后生成字段的个数，如图6-54所示。



图6-54　自定义拆分设置

本节以“班”字段为例，介绍Tableau默认的“拆分”功能。右键单击维度窗口中的“班”字段，选择“变化”➤“拆分”，此时，在维度窗口中，生成了新的字段“班 – 拆分1”，通过对比可发现，Tableau自动将字段中的数字拆分出来，生成了数字型的字段，如图6-55所示。



图6-55　使用默认拆分生成新的字段

6.7　参考线及参考区间

Tableau在分析中嵌入了参考线（Reference Line）、参考区间（Reference Band）、分布区间（Distribution Band）和盒须图（Box Plot），来标记轴上的特定值或区域。

1. 参考线

参考线即在轴上添加一条线，用来标记某个常量或计算值位置。该计算值可基于指定的字段或参数生成，常用的有该轴的平均值、最小值、最大值等。参考线可基于表、区或单元格进行设置。以设置平均值为例，基于“表”“区”“单元格”的设置效果如图6-56所示。

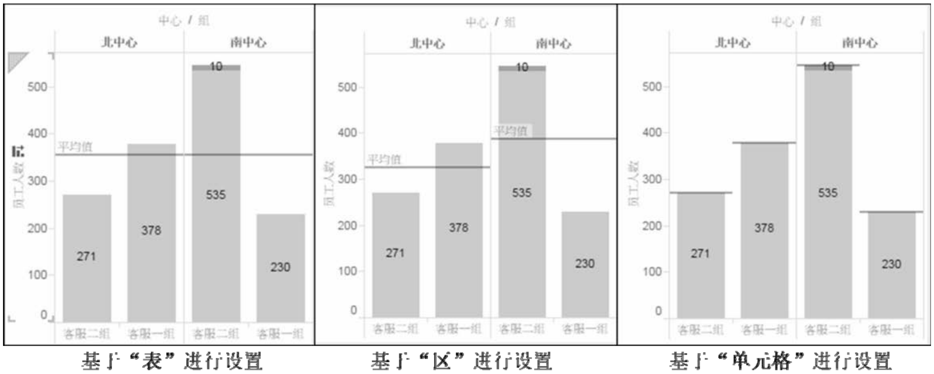


图6-56　“人工服务接听量”平均值参考线

2. 参考区间

参考区间是指在轴上添加一个区间，用来标记某个范围，将视图标记之后，轴上两个常量或计算值之间的区域显示为阴影，如图6-57所示。

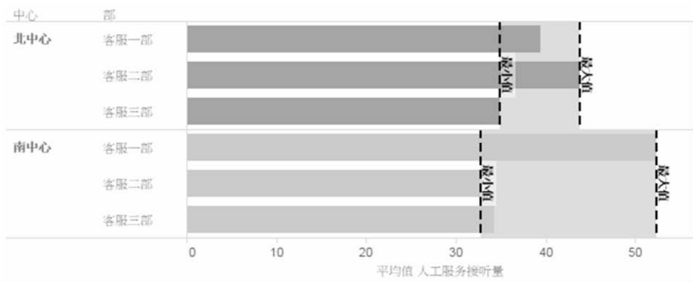


图6-57　“人工服务接听量”最小值 – 最大值每区参考区间

3. 分布区间

分布区间是指通过添加阴影梯度或组合参考线来指示沿轴的数值分布情况，如图6-58所示。该分布可以通过置信区间、百分比、百分位、分位数或标准差来定义。

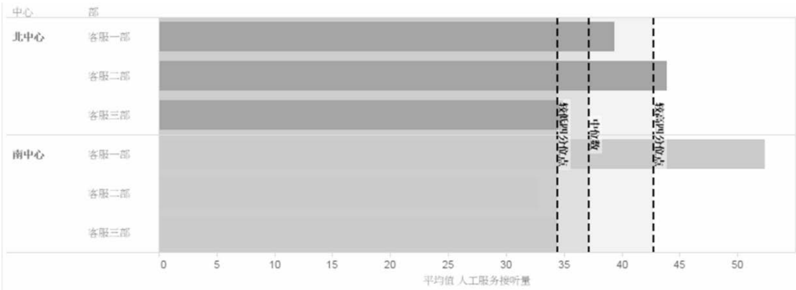


图6-58　“人工服务接听量”四分位分布区间

4. 盒须图

盒须图是用来显示一组数据分散情况的统计图，能提供有关数据位置和分散情况的关键信息。如图6-59所示，从左到右的五条线分别表示最小值、下四分位数、中位数、上四分位数和最大值。Tableau 提供了多种样式，并且允许配置“须”线的位置和其他信息，具体内容请参见7.2节。

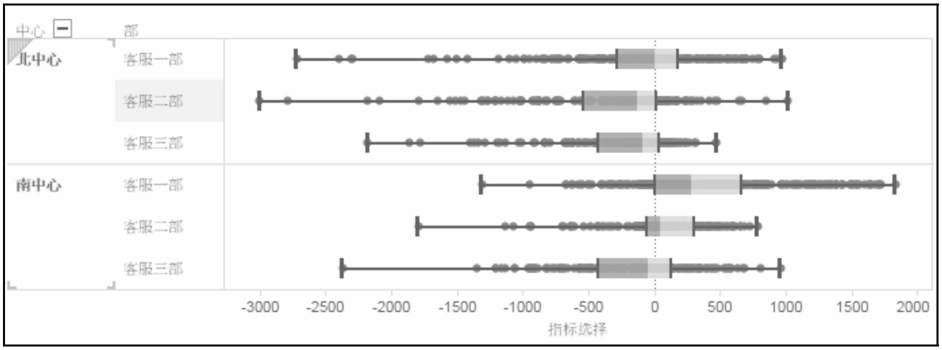


图6-59　盒须图示例

在分析窗口中，Tableau将常用的功能在“汇总”中进行了列示，包括常量线、平均线、含四分位点的中值、盒须图和合计。

由于分布区间的创建和使用与参考线、参考区间非常相似，且盒须图在7.2节有详细介绍，因此本节将重点介绍参考线及参考区间的创建和使用方法。

6.7.1　创建参考线及参考区间

在创建参考线之前，需要构建基本视图，首先构建一个平均“人工服务接听量”条形图，如图6-60所示。

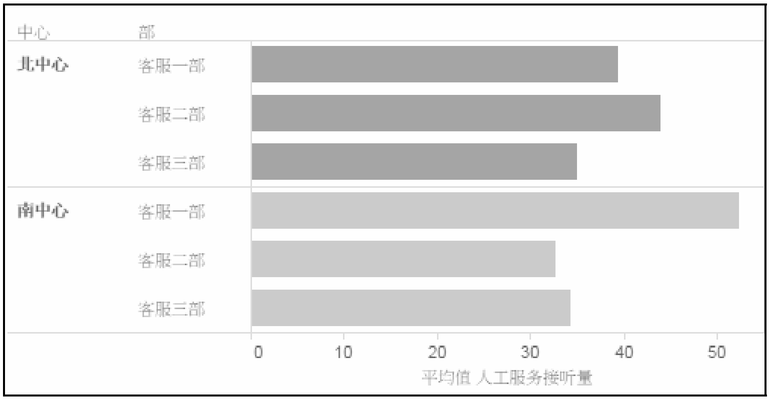


图6-60　各中心、部门平均“人工服务接听量”条形图

1. 参考线：添加“人工服务接听量”平均值参考线

为掌握各个中心的“人工服务接听量”平均值，以对比各个部的接听量多少，需在各中心添加中心内的“人工服务接听量”平均值，添加方法如图6-61所示，拖放“分析窗口”➤“汇总”下的“平均值”到视图中，并在弹出的对话框中选择“区”即可。通过添加参考线可看出，南中心每人每天的接听量大于北中心，并且南中心客服一部的平均接听量远远大于中心平均值。

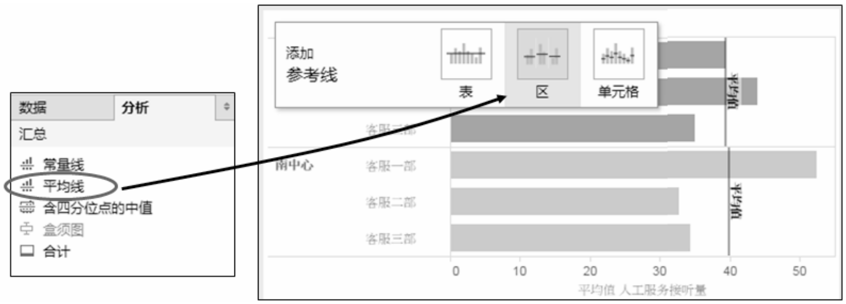


图6-61　各中心、部门平均“人工服务接听量”条形图

2. 参考线：添加“服务满意评价满意数”平均值参考线

除快速添加常用参考线外，Tableau有两种方式可添加参考线：①拖放分析窗口处的“参考线”到视图中；②右键单击对应的轴，在弹出的菜单中选择“添加参考线”，这两种方式的效果是一致的。

在对比各个中心的工作时，除接听量为考核指标外，每天服务评价的满意程度也是重要指标，因此添加“服务评价满意数”为参考线，对比各中心、各部的接听量和服务满意数的差别。下面以拖放“参考线”到视图为例，介绍参考线的添加和设置方式。

(1) 为使用“服务评价满意数”创建参考线，首先拖放该字段到“详细信息”中。

(2) 直接拖放分析窗口中的“参考线”到视图中，在弹出的对话框中设置参考线的“范围”为“每单元格”，将线的值选为“平均值（服务评价满意数）”，将标签设为“值”（标签为显示在视图中的内容，包括“无”“值”“计算”和“自定义”4种模式），样式为默认选项，如图6-62所示。

创建后的视图如图6-63所示。添加参考线之后，可右键单击该参考线，在弹出的菜单中选择相应的选项进行设置，其中单击“平均值”可修改参考线的计算方式，包括平均值、最小值和最大值等；单击“编辑”，弹出参考线的设置界面，可对参考线重新进行设置；单击“设置格式”，在边条上显示参考线、标签的格式设置选项；单击“移除”可直接删除参考线。



图6-62　创建“服务评价满意数”平均值参考线

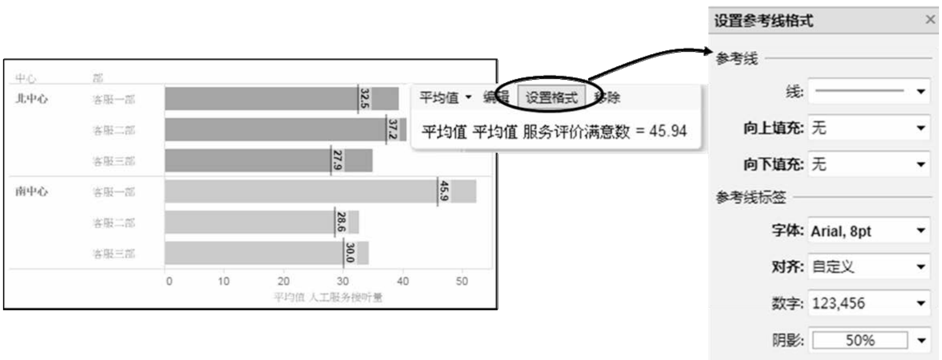


图6-63　添加“服务评价满意数”后的视图

6.7.2　创建参考区间

为对比中心各部的人工服务接听量的大小，可添加中心内各部的接听量最小值和最大值的参考区间。Tableau有以下两种方式添加参考区间。

- 方式1：直接拖放分析窗口中的“参考区间”到视图中，在弹出的对话框中设置参考线的“范围”为“每区”，将区间开始的值选为“平均值（人工服务接听量）”的“最小值”，区间结束的值选为“平均值（人工服务接听量）”的“最大值”，样式为默认选项，生成的视图如图6-64所示。



图6-64　对每区的平均“人工服务接听量”创建最小值–最大值参考区间

- 方式2：右键单击X轴，选择“添加参考线”，在弹出的对话框中，选择“区间”，其他设置同方式1。

第 7 章 高级可视化分析

本章将综合运用上一章介绍的Tableau高级特性来创建高级视图，包括帕累托图、盒须图、瀑布图、范围–线图、倾斜图、网络图和雷达图的创建和使用。

7.1 帕累托图

帕累托（Pareto）图是按照一定的类别根据数据计算出其分类所占的比例，用从高到低的顺序排列成矩形，同时展示比例累积和的图形，主要用于分析导致结果的主要因素。帕累托图与帕累托法则（又称为“二八原理”，即80%的结果是20%的原因造成的）一脉相承，通过图形体现两点重要信息：“至关重要的极少数”和“微不足道的大多数”。

本节以某企业的物资采购金额数据为例创建一个帕累托图，帮助大家迅速发现隐藏在数据中的重要信息。该图由柱形图、折线图和参数控件3部分组成。其中，横轴为供应商数量占总供应商数量的累计比例，柱形图显示由大到小的各供应商的应付金额情况，折线图显示金额累计百分比沿着横轴的变化情况，参数控件帮助快速定位参考线的位置。如图7-1所示，当参数调整到80%时，这时X轴的参考线为13.7%，表示该企业80%的物资采购的应付金额集中于13.7%的供应商，因此这13.7%的供应商值得我们关注。

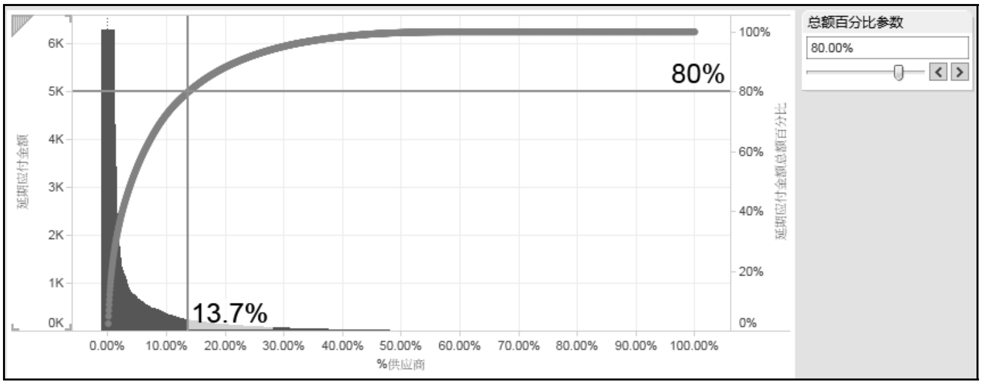


图7-1　帕累托图

说明 供应商百分比化：将供应商对应于0%~100%的区间。比如有200个供应商，它们按照一定顺序排列，若某供应商处于第60位，则其百分比为60 / 200 * 100% = 30%（代表着按照该顺序降序排序的前30%的电厂）。

1. 数据准备

导入数据“物资采购金额.xls”（见附录D.8），完成本案例所需的字段为“供应商名称”和“应付金额”。

2. 创建累计百分比图

利用表计算函数RUNNING_SUM和TOTAL创建计算字段“应付金额总额百分比”，该字段可求出某供应商之前（按照横轴从左往右）的所有供应商应付金额总和占总应付金额的百分比（见图7-2）。

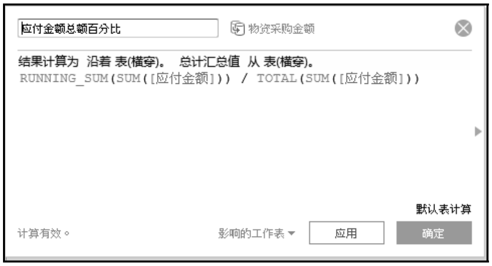


图7-2 应付金额总额百分比

将“供应商名称”拖放到列功能区，将“应付金额总额百分比”拖放到行功能区，计算依据选择“供应商名称”。为显示所有供应商，选择视图为“适合宽度”（见图7-3）。

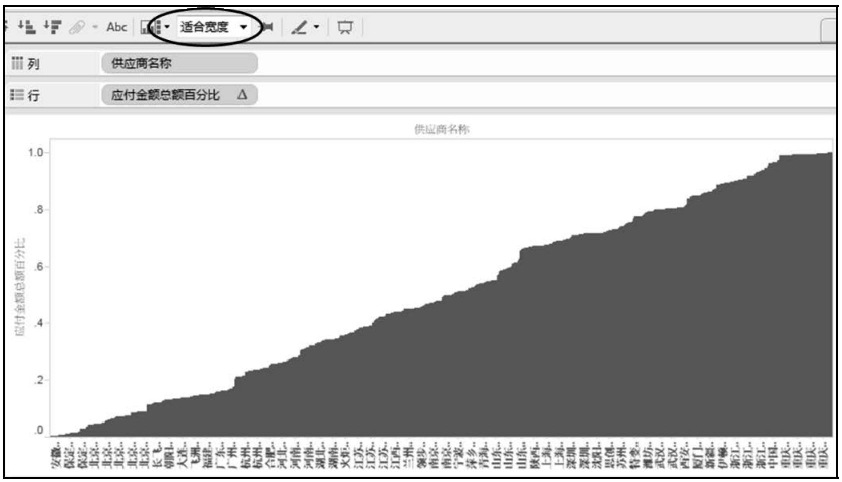


图7-3 创建各供应商的应付金额百分比视图

对“供应商名称”进行排序，按照 “应付金额”字段的“总计”值进行降序排列。在“标记”卡中将图形选择为“线图”，即完成累计百分比图（见图7-4）。

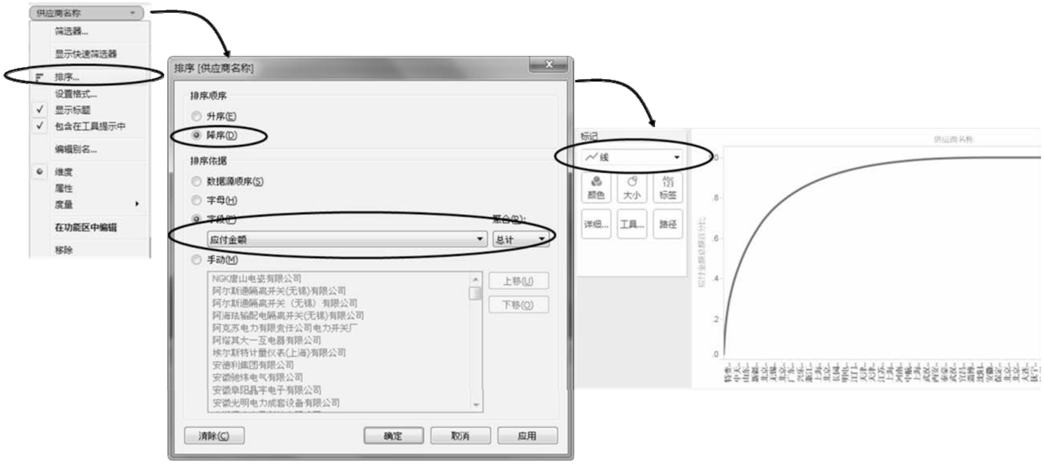


图7-4　应付金额总额百分比图

3. 创建柱形图

在图7-4的基础上将字段“应付金额”拖放到行功能区，调整“标记”卡中“应付金额”为条形图。请在字段“应付金额”上单击右键，在弹出的下拉列表中选择“双轴”，这时两个图形将按双轴合并。此时请调整“应付金额”和“应付金额总额百分比”两个字段的左右顺序，使“应付金额”显示在左轴，生成图7-5。

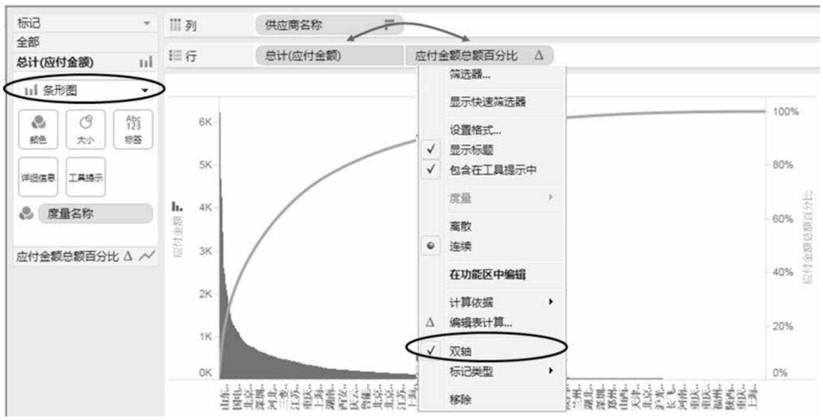


图7-5　应付金额柱形图

此刻横轴的供应商较多，为了更好地表示分布，我们将横轴转换为供应商总数量的百分比进行展示。

创建计算字段“%供应商”，公式为index()/size()，指该供应商之前（排序后从左往右数）的供应商数量占总供应商数的百分比。将创建的字段拖放到列功能区，并单击右键，在弹出的下拉列表中选择“计算依据”➤“供应商名称”，将原有字段“供应商名称”拖放到“标记”卡上的“全部”页签内，见图7-6。

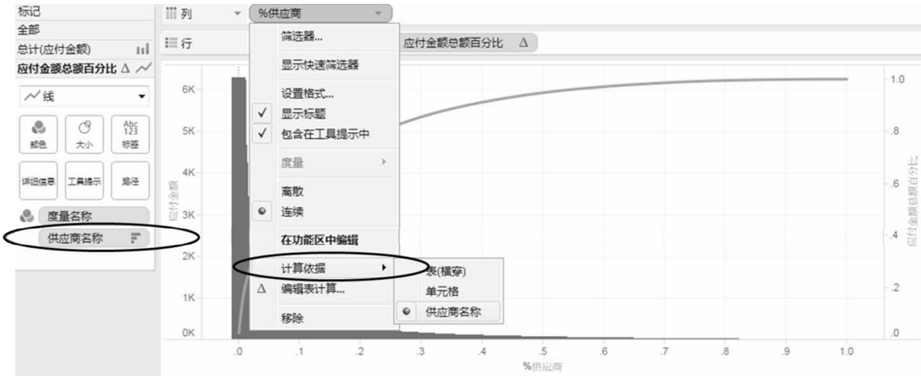


图7-6　帕累托图

注意　创建的计算字段默认显示为小数，需要设置坐标轴的展示数字格式为百分比。

4. 创建动态参数

图7-6已生成了基本的帕累托图，但该图还不能直观地显示“至关重要的极少数”是多少。为了快速获取这些信息，现在介绍利用参数、自定义计算字段和参考线综合实现的方式。

创建参数“总量百分比参数”作为应付金额总额百分比（线图）纵轴的参考线，其数据类型为浮点数，允许选择的最小值为0，最大值为1，步长为0.01，具体设置如图7-7所示。



图7-7　创建总额百分比参数

为了让累计百分比图的横轴参考线和纵轴参考线的交点落在累计百分比图上，我们需要创建一个新的字段作为横轴参考线的值的依据。

创建字段“横轴参考线（百分比）”，计算公式为IF [应付金额总额百分比]<=[总额百分比参数] THEN [%供应商] ELSE NULL END，参见图7-8。

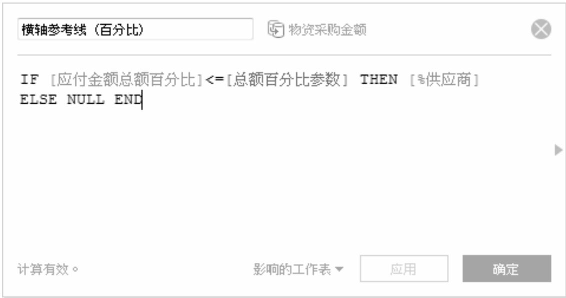


图7-8　创建横轴参考线（百分比）字段

图7-9为“应付金额总额百分比”纵轴添加参考线，其值为“总额百分比参数”，标签为“值”，并设置线的形状、颜色等；为横轴添加参考线，其值为“横轴参考线（百分比）”的“最大值”，同样设置其标签为“值”，并调整线的形状和颜色，最终生成如图7-1所示的效果，当调整参数“总额百分比参数”时，“横轴参考线（百分比）”动态变化，两个参考线的交点始终落在线图上。



图7-9　为纵轴、横轴添加参考线

7.2 盒须图

盒须图又叫箱线图，是一种常用的统计图形，用以显示数据的位置、分散程度、异常值等。箱线图主要包括6个统计量：下限、第一四分位数、中位数、第三四分位数、上限和异常值（参见图7-10）。

- **中位数**：数据按照大小顺序排列，处于中间位置，即总观测数50%的数据。
- **第一四分位数、第三四分位数**：数据按照大小顺序排列，处于总观测数25%位置的数据为第一分位数，处于总观测数75%位置的数据为第三分位数。四分位全距是第三分位数与第一分位数之差，简称IQR。
- **上限、下限**：Tableau可设置上限和下限的计算方式，一般上限是第三分位数与1.5倍的IQR之和的范围之内最远的点，下限是第一分位数与1.5倍的IQR之差的范围内最远的点。也可直接设置上限为最大值，设置下限为最小值。
- **异常值**：在上限和下限之外的数据。

一般来说，上限与第三四分位数之间以及下限与第一四分位数之间的形状称为**须状**。

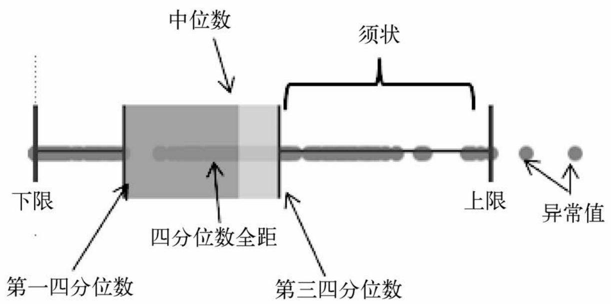


图7-10 盒须图

通过绘制盒须图，观测数据在同类群体中的位置，可以知道哪些表现好，哪些表现差；比较四分位全距及线段的长短，可以看出哪些群体分散，哪些群体更集中。

这里以坐席接听统计数据为例，对南北中心各部人员平均呼入通话时长进行分析做出箱线图，见图7-11。

7

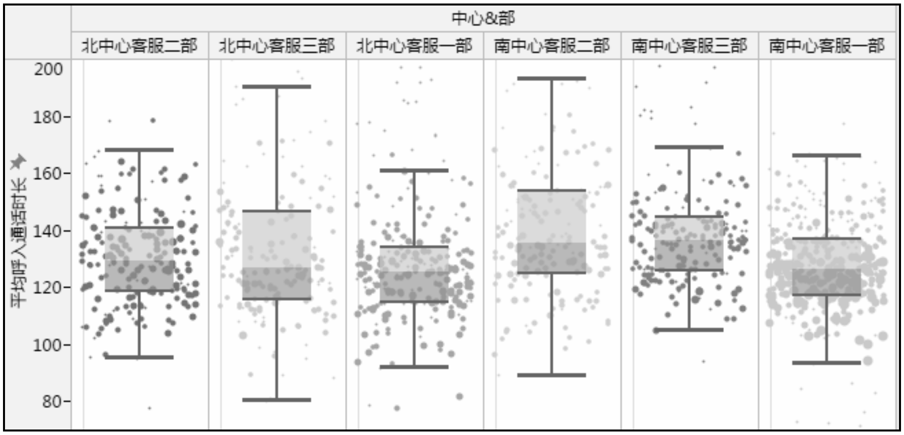


图7-11　箱线图

7.2.1　基础应用

本节基于实例介绍如何创建盒须图，以便分析数据的中心位置及离散情况。

1. 数据准备

导入坐席接听统计数据，完成该案例需要的维度字段有“中心”和“部”，度量字段包括“呼入通话时长（秒）”和“人工服务接听量”。

2. 创建箱线图

- ***步骤1：创建所需字段***

(1) 中心分为南中心和北中心，每个中心各有3个部，因此需要将字段“中心”和“部”进行合并，以创建字段“中心&部”，其计算公式见图7-12（左），或直接使用创建合并字段功能。

(2) 为分析每个人接听电话的平均通话时长，我们需要创建字段“平均呼入通话时长”，计算原理为一个人一个月总通话时长除以总接听量，计算公式见图7-12（右）。



图7-12　创建所需字段

● **步骤2：生成基本视图**

(1) 将创建好的字段“平均呼入通话时长”和“中心&部”分别拖放到行功能区和列功能区。

(2) 拖放“工号”到“标记”卡，图形选择“圆”视图。这时视图中每一个圆点即代表一个工号，字段“平均呼入通话时长”会对每一个工号计算其平均通话时长（参见图7-13）。

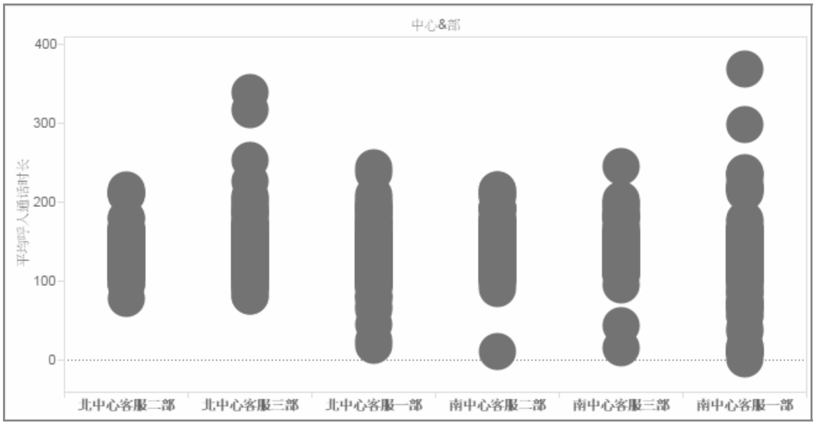


图7-13　生成基本视图

● **步骤3：创建箱线图**

(1) Tableau有两种创建盒须图的方式：单击“智能显示”➤“盒须图”完成盒须图视图；拖放“分析”窗口中的“盒须图”到视图中。

(2) 对第一种方式创建的盒须图，右键单击纵轴，选择“编辑参考线”，在弹出的对话框中设置盒须图的格式，见图7-14；或直接单击盒须图，选择“编辑”进行设置。第二种方式创建的盒须图会直接弹出设置界面。请设置盒须图的样式，包括样式、填充、边界等的格式。

7

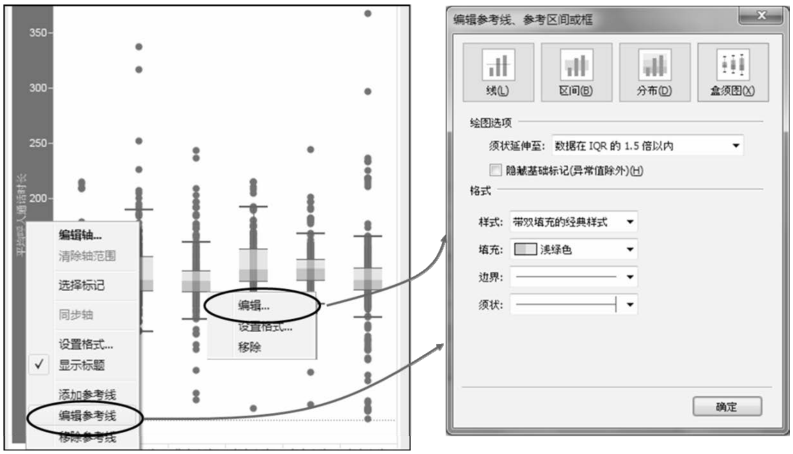


图7-14　盒须图格式设置

(3) 设置成功后，单击“确定”，生成盒须图，如图7-15所示。

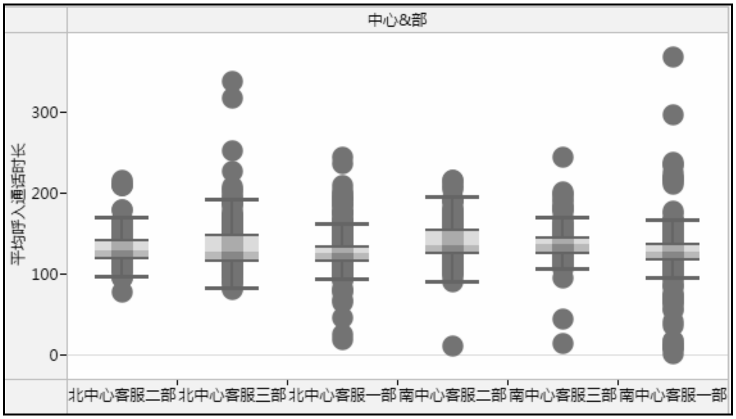


图7-15　生成盒须图

7.2.2　图形延伸

如图7-15所示，所有的点都落在了一条垂直线上，一个点代表一个工号，由于工号较多，很多点都是重叠覆盖的，不能直观地展示各部之间人员数量的比较，也无法直观显示各部内员工的分布。这里介绍将点水平铺开的方法，最终生成效果如图7-16所示。

(1) 创建自定义计算字段“将点散开”，计算公式为`index()%30`。

(2) 将其拖放到列功能区“中心&部”的右边，设置“计算依据”为“工号”，各个圆点即水平展开，展开幅度为30。我们可调整“将点散开”的公式来调整散开的幅度。

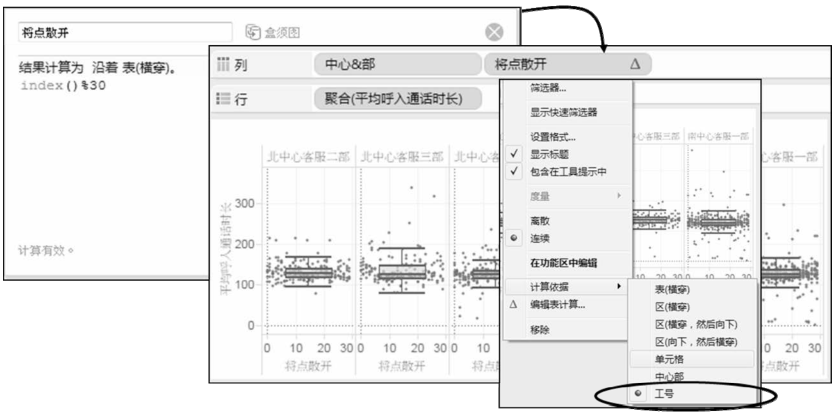


图7-16　设置将点散开效果

(3) 为了分析平均呼入通话时长的异常点问题所在，我们将“人工服务接听量”拖放到选项卡

中的“大小”，同时为了使图形更美观，将“中心&部”拖放到“颜色”，生成结果如图7-17所示。

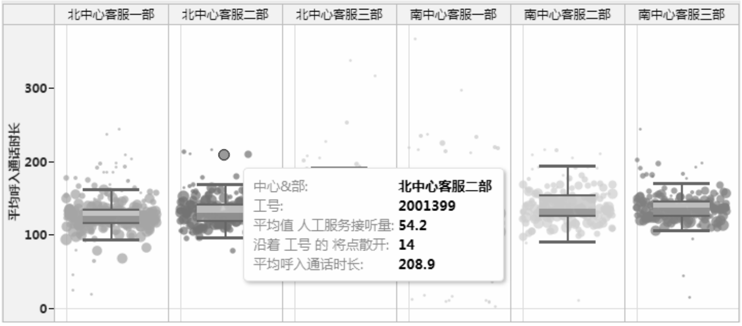


图7-17 设置将点散开效果（另见彩插图7-17）

通过图形分析可看出，平均呼入通话时长异常的点（须外的点），人工服务接听量普遍较少，不具有分析价值。而我们应重点关注通话时长异常并且接听量也较多的员工，如北中心客服中心二部的2001399号员工，平均每天接听54.2个电话，每个电话均长达208.9秒，值得重点关注。

7.3 瀑布图

瀑布图是数据可视化分析中常见的一种图形，采用绝对值与相对值结合的方式，适用于表达数个特定数值之间的数量变化关系。对于一系列具有累计性质的正值/负值具有很好的展示功能，既可以辅助理解数据的大小，又能直观地展示出数据的增减变化，反映数据在不同时期或受不同因素的影响结果。

以某地区各单位预转资对折旧费的影响为例，使用“影响折旧费”数据创建如图7-18所示的瀑布图，从左往右直观展示了各单位预转资对折旧费的影响。其中从左数第1个到第7个柱形代表对折旧费的影响为正值（即超时预转资），第8个到第9个柱形代表对折旧费的影响为负值（即提前预转资），最右端的“总计”为影响折旧费的总计值。

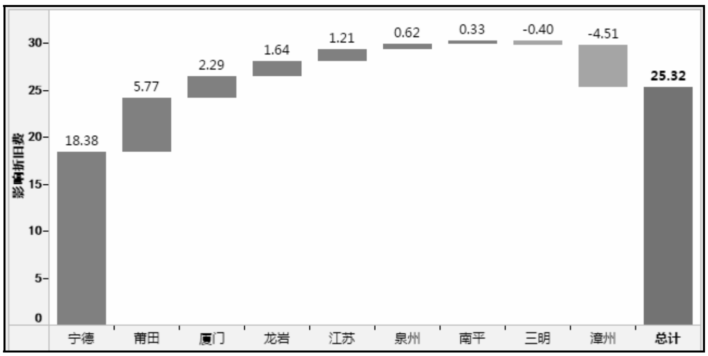


图7-18 瀑布图（另见彩插图7-18）

7.3.1 基础应用

本节基于实例介绍如何创建瀑布图，以有效展示数值间的变化关系。

1. 数据准备

导入影响折旧费数据，完成折旧费影响分析需要数据表中的“单位名称”和“影响折旧费”这两个字段。

2. 创建瀑布图

在Tableau中，瀑布图是由甘特图生成的。首先画出各单位影响折旧费的累计柱形图，再将柱形图转换为甘特图，最后再次利用折旧费影响值定义甘特图的大小即可。下面将分步骤介绍瀑布图的完成方法。

- **步骤1：完成甘特图**

(1) 将“单位名称”和“影响折旧费”分别拖放到列功能区和行功能区，并对“影响折旧费”添加快速表计算“汇总”，计算依据为“单位名称”，生成柱形图。

(2) 在“标记”卡中选择图形为甘特图（见图7-19）。

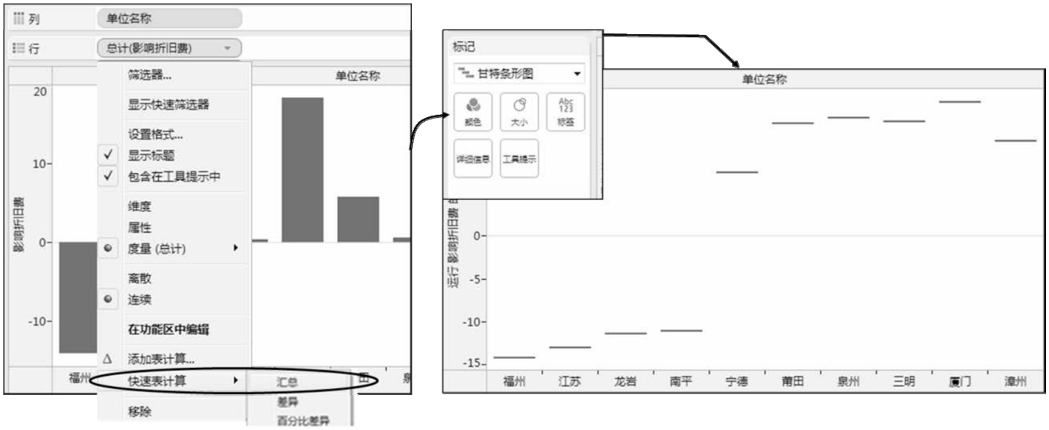


图7-19　完成甘特图

- **步骤2：完成基本的瀑布图**

(1) 创建计算字段“–影响折旧费”，定义为“影响折旧费”的负值，将其拖放到“标记”卡中的“大小”里，值的大小显示为柱子的高低，值的正负显示为不同方向，即以甘特图的位置为基准。若“–影响折旧费（万元）”为正，则方向向上；若“–影响折旧费”为负，则方向向下，如图7-20所示。

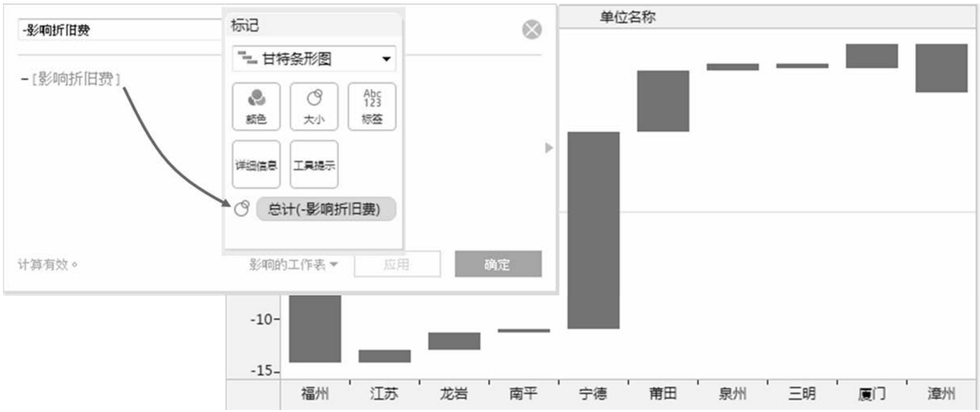


图7-20　完成甘特图

(2) 对各单位进行排序，按照“影响折旧费”字段的值从大到小降序排列。右键“影响折旧费”或左键“影响折旧费（万元）”右侧小三角形，选择“排序”，在弹出对话框中选择“降序”，按字段“影响折旧费”的总计作为排序依据，单击“确定”（见图7-21）。

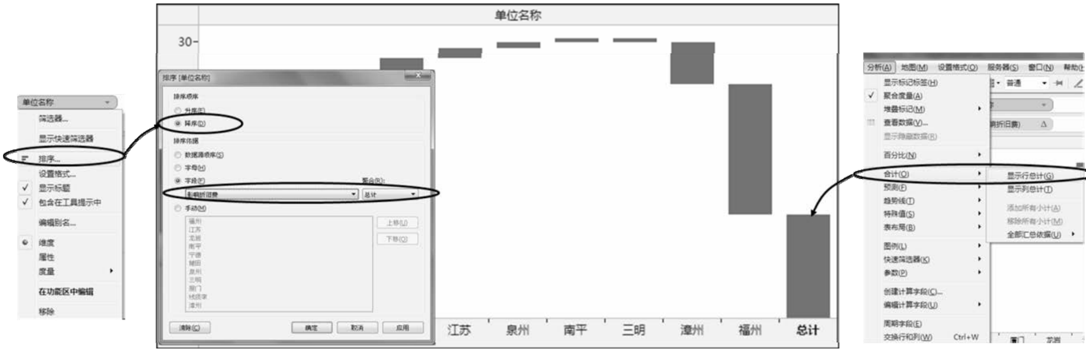


图7-21　完成基本的瀑布图

(3) 在菜单栏选择“分析“➤“合计”➤“显示行总计”，此时生成各单位影响因素的总和，即折旧费的变化。

● **步骤3：根据字段“影响折旧费”值的正负定义单位的不同颜色**

(1) 将字段“影响折旧费”拖放到“标记”卡中的“颜色”里，在弹出的颜色图例中“编辑颜色”。

(2) 在弹出的颜色编辑框中，将“颜色渐变”选择为二阶，单击“高级”，为中心选择“0”，这时就以0为划分两阶的依据，如图7-22所示。

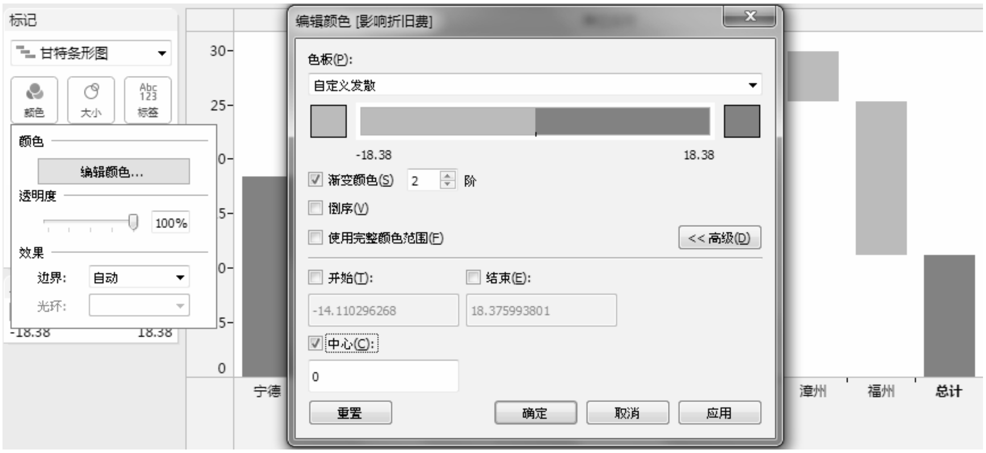


图7-22　定义单位的不同颜色

(3) 若需要“总计”与各单位的柱形图颜色有所区别，可创建计算字段“影响类别”（其计算公式为IF [影响折旧费]>=0 THEN '正向' ELSE'负向' END）；然后拖放其到“标记”卡中的“颜色”中，修改其为“属性”（见图7-23），你可通过编辑颜色修改各字段的颜色。

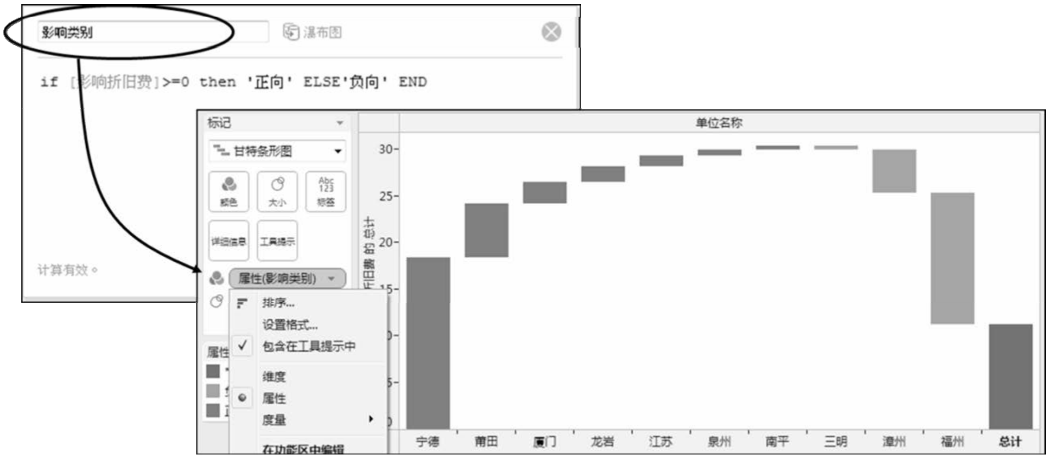


图7-23　自定义“总计”颜色

7.3.2　图形延伸

Tableau做瀑布图的过程可延伸出“变化排序”的图形，可展示多维信息。以展示不同单位2013年和2014年售电量的变化为例，如图7-24所示虚线表示各单位2013年的售电量，柱形大小表示2014年减去2013年的差值以表示变化，其中绿色表示上升，红色表示下降，柱子的末端为各单位2014年的售电量。通过一个图形我们可以直观地看出各单位2014年售电量的排名、2013年的售电

量情况以及各单位的同比变化情况。

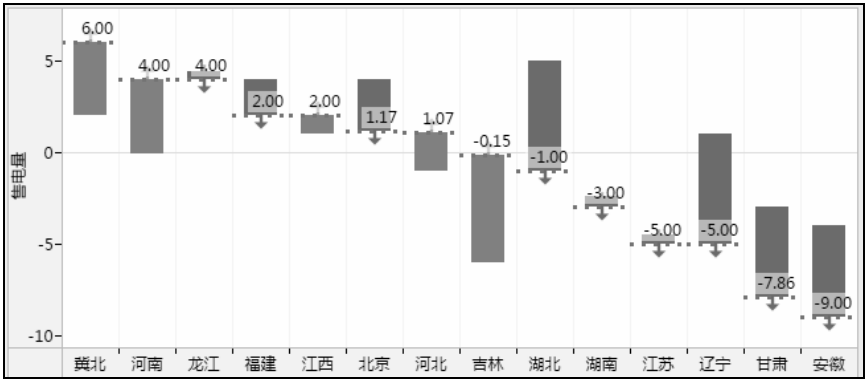


图7-24 “变化排序”图（另见彩插图7-24）

“变化排序”图的作图方式与瀑布图类似，以“2014年各省市售电量明细表”数据为例，主要步骤如下。

(1) 将“省市”和“同期值”分别拖放到列功能区和行功能区，筛选“统计周期”为2014年，并将图形修改为“甘特条形图”。

(2) 创建计算字段“同比变化量”，计算公式为“[当期值]-[同期值]”；创建计算字段“增加or减少”，计算公式为IF ([当期值]-[同期值])>=0 THEN '增加' ELSE '减少' END。拖放“同比变化量”到“大小”，“增加or减少”到“颜色”，如图7-25所示。

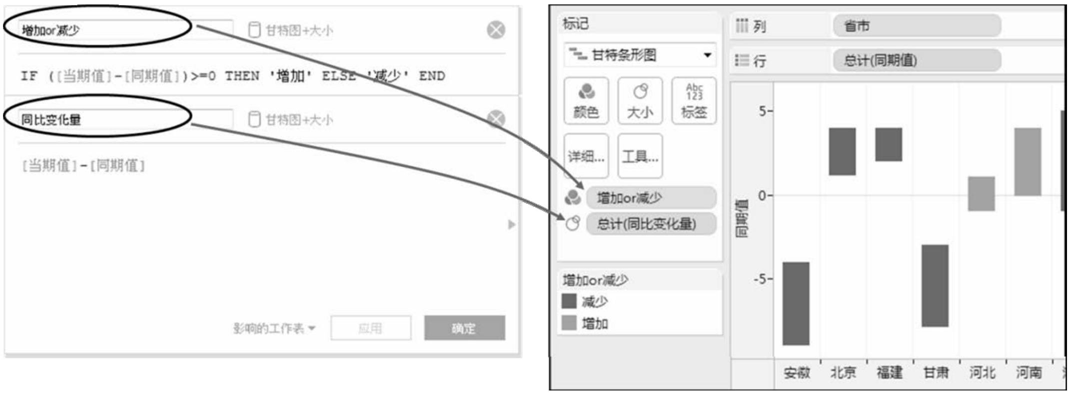


图7-25 创建计算字段，生成基本视图

(3) 拖放字段“当期值”到行功能区，设置图形为“形状”，拖放“增加or减少”到“标记”卡中的“颜色”和“大小”中，单击双轴，并同步轴，如图7-26所示。

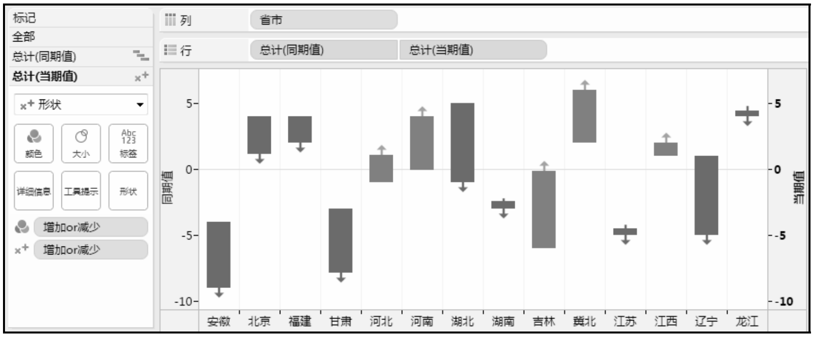


图7-26　创建双轴视图

(4) 为“当期值”添加参考线，针对每个单元格设置参考线为“当期值”，标签为“值”，格式为虚线。

(5) 单击省市进行排序，按照“当期值”降序排列。

(6) 单击“当期值”的纵轴，选择“显示标题”，取消该轴的显示；单击“同期值”的纵轴进行编辑，修改标题为“售电量”，如图7-27所示，完成以上 6 步操作后，最终生成效果如图7-24所示。



图7-27　设置轴的显示效果

7.4　范围—线图

折线图是常见的可视化图形，但其包含的信息量少，在做深化分析时往往需要进行优化。

以坐席接听统计数据为例，某员工某月各日的接听量趋势如图7-28中的单一折线图所示（图中只展示了上班的日期），但该图并不能反映这名员工在整个员工群体中所处的位置。**范围–线图**（range-line chart）将群体数据的部分统计特征（如均值、中位数、最大值、最小值、分位数等）展示在图形中，既可以说明群体的信息，也可以展示个体的信息，还可以比较个体与整体的相对

位置关系，展示的信息更为丰富。以图7-28为例，选取群体的最大值、最小值和均值在图中展示，可看出该员工基本日接听量都在平均线以上，但并不是最优秀的，因为与最好的员工相比（最大值）还有一定距离。

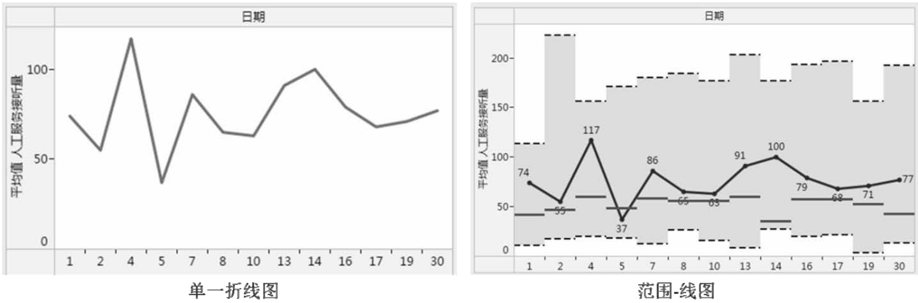


图7-28　单一折线图和范围–线图对比

1. 数据准备

范围–线图中的最大值、最小值和均值是依据字段“人工服务接听量”添加的参考区间和参考线。但因为折线展示的字段是单个人的“人工服务接听量”，Tableau无法对同一个字段既展示整体又展示个体，因此我们需要改变思路，调整数据源以实现展示效果。

在数据中为每一行数据添加三列，这三列分别为：每日所有员工的人工服务“接听量最大值”“接听量平均值”和“接听量最小值”。数据源改变后如图7-29所示。



F	I	J	AC	AD	AE
日期	工号	人工服务接	接听量最大值	接听量平均值	接听量最小值
1/1/2014	2000241	66	113	42	11
1/1/2014	2000261	22	113	42	11
1/1/2014	2000264	43	113	42	11
1/1/2014	2000271	22	[illegible]	[illegible]	11
[illegible]	[illegible]	14	[illegible]	[illegible]	11
[illegible]	[illegible]	51	[illegible]	[illegible]	17
[illegible]	[illegible]	8	[illegible]	[illegible]	17
[illegible]	[illegible]	50	[illegible]	[illegible]	17
1/2/2014	2000943	12	223	47	17
1/2/2014	2000956	26	223	47	17

图7-29　数据准备

2. 创建范围–线图

● **步骤1：生成基本视图**

(1) 将“日期”和“人工服务接听量”拖放到行功能区和列功能区，其中“日期”显示形式为“离散”的“日”。

(2) 将“接听量平均值”“接听量最大值”和“接听量最小值”均拖放到“标记”卡中的“详细信息”中，聚合方式为“平均值”。

(3) 将“工号”拖放到“筛选器”卡中，并任意选择一个工号。如图7-30所示。

7

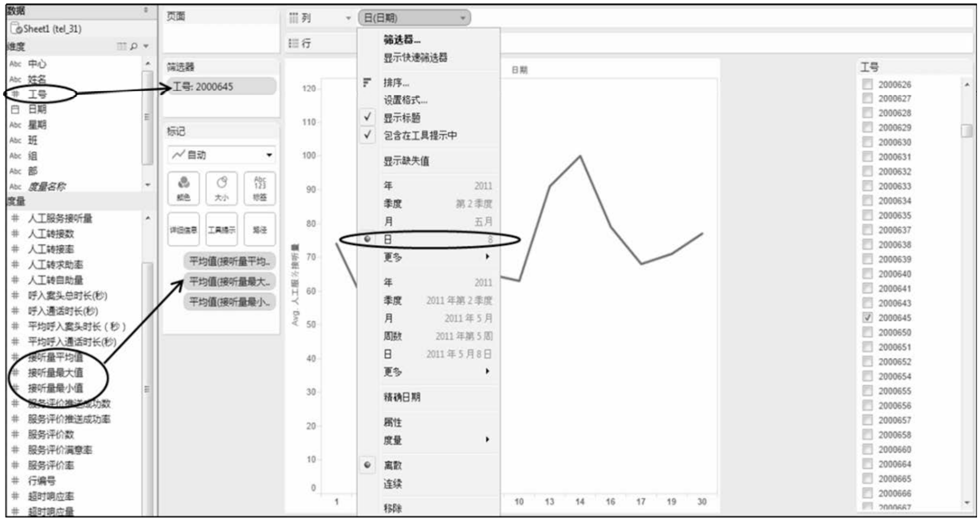


图7-30　拖放相关字段

- **步骤2：添加参考线和参考区间**

(1) 添加参考区间：选择“区间”，为范围选择“每单元格”，为最小值选择“平均值（接听量最小值）”，为最大值选择“平均值（接听量最大值）”。

(2) 添加参考线：选择“线”，为范围选择“每单元格”，为线选择“平均值（接听量平均值）”参见图7-31。



图7-31　添加参考线及参考区间

完成以上操作后，最终生成图形如图7-28（右）所示。

7.5　倾斜图

倾斜图（slopegraph）可以展示单指标不同时期的变化，既能展示值的大小变化，也能同时展示排名的变化。以“全员劳动生产率”指标为例看各单位当期值和同期值的变化情况，做出的倾斜图如图7-32所示。分别对同期和当期的各单位按照累计值大小进行排名，名次显示在“单位”标签的左边，“单位”标签的右边为累计值。线的倾斜方向显示着名次的上升或下降，颜色表明累计值的增加或减少（绿色表示增加，红色表示减少），线的粗细表示累计值变化量的绝对值。从图中可看出，浙江、天津排名增加较多，同时量的增幅也是最大的，而像吉林、黑龙江等虽然是在增加，但名次反而降低，说明相对表现在退步。

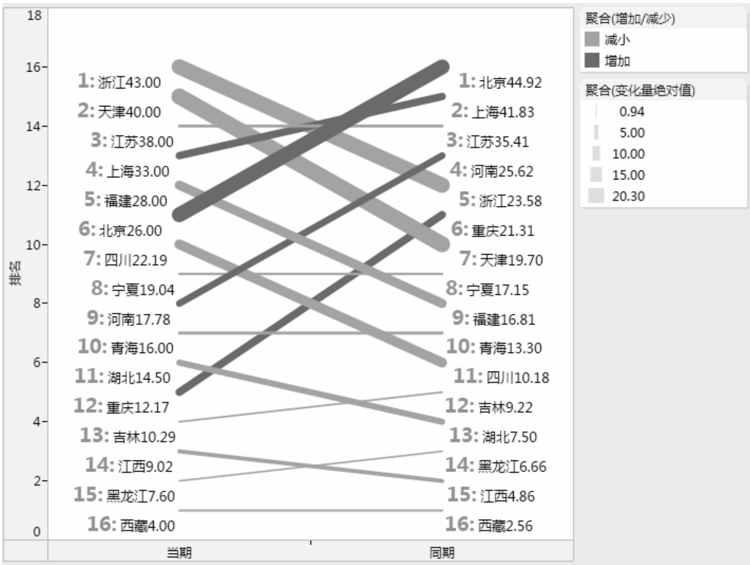


图7-32　倾斜图（另见彩插图7-32）

创建倾斜图需要运用Tableau自定义计算字段、表计算等高级特性，同时标记卡、格式设置等基本功能也使用较多。

1. 数据准备

为实现各单位的连线效果，我们需要调整基础数据，将“当期累计值”和“同期累计值”调整为一列，新增字段“期间”以区分当期和同期。改变后的数据结构如图7-33（右）所示。

A	B	C	D
指标名称	单位	当期累计值	同期累计值
全员劳动生产率	上海	33	41.83
全员劳动生产率	北京	26	44.92
全员劳动生产率	吉林	10.29	9.22
全员劳动生产率	四川	22.19	10.18
全员劳动生产率	天津	40	19.7
全员劳动生产率	宁夏	19.04	17.15
全员劳动生产率	江苏	38	35.41
全员劳动生产率	江西	9.02	4.86
全员劳动生产率	河南	17.78	25.62
全员劳动生产率	浙江	43	23.58
全员劳动生产率	湖北	14.5	7.5
全员劳动生产率	福建	28	16.81
全员劳动生产率	西藏	4	2.56
全员劳动生产率	重庆	12.17	21.31
全员劳动生产率	青海	16	13.3
全员劳动生产率	黑龙江	7.6	6.66

→

A	B	C	D
指标名称	单位	期间	累计值
全员劳动生产率	上海	当期	33
全员劳动生产率	北京	当期	26
全员劳动生产率	吉林	当期	10.29
全员劳动生产率	四川	当期	22.19
全员劳动生产率	天津	当期	40
全员劳动生产率	宁夏	当期	19.04
全员劳动生产率	江苏	当期	38
全员劳动生产率	江西	当期	9.02
全员劳动生产率	河南	当期	17.78
全员劳动生产率	浙江	当期	43
全员劳动生产率	湖北	当期	14.5
全员劳动生产率	福建	当期	28
全员劳动生产率	西藏	当期	4
全员劳动生产率	重庆	当期	12.17
全员劳动生产率	青海	当期	16
全员劳动生产率	黑龙江	当期	7.6
全员劳动生产率	上海	同期	41.83
全员劳动生产率	北京	同期	44.92

图7-33　数据准备

2. 创建斜线图

- ***步骤1：搭建基本框架***

(1) 创建“排名”字段，公式为index()，并将其拖放到行功能区。

(2) 将字段“期间”拖放到列功能区，将字段“单位”拖放到“标记”卡中。

(3) 对字段“排名”定义表计算，选择“根据以下因素计算”➤“高级”，在对话框中将“寻址”设为“期间”和“单位”（注意“期间”在上）；设置排序依据为“累计值”，升序排列。回到表计算中设置 “重新启动间隔”➤“期间”。最终下面的表计算说明显示“结果计算为沿着单位（按累计值总计的升序排序）对于每个期间”。结果如图7-34所示。

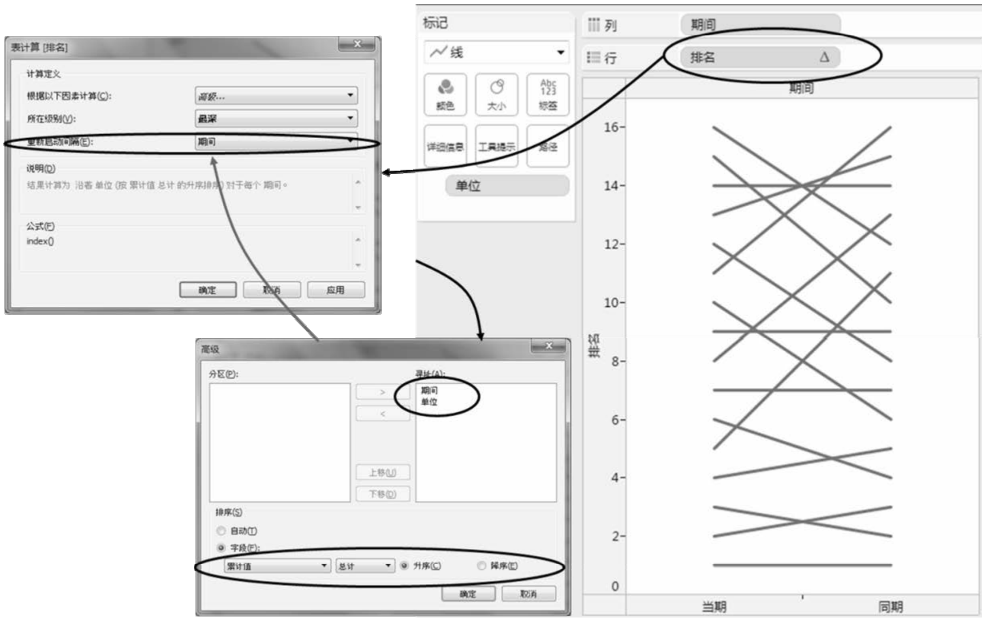


图7-34　搭建基本框架

● **步骤2：添加标签**

(1) 从图7-34中可以看出，表现最好的单位显示排名为16，因此我们创建一个新的排名字段“排名(2)”，计算公式为`17-index()`。

(2) 将字段“单位”“累计值”和“排名(2)”分别拖放到“标签”中。

(3) 设置“排名(2)”的表计算，与“排名”的设置方式一致，最终下面的表计算说明显示“结果计算为沿着单位（按累计值总计的升序排序）对于每个期间”，这样一来表现最好的单位显示排名为1。

(4) 单击“标记”卡中的“标签”，在弹出的对话框中选择“文本”，对3个标签字段排版、格式进行设置，突出“排名(2)”；在选择对齐方式时，为水平选择“自动”，为垂直选择“中部”；“标签标记”➤“线末端”，“选项”全部勾选，包括“允许标签覆盖其他标记”“线首标签”和“线尾标签”。结果如图7-35所示。

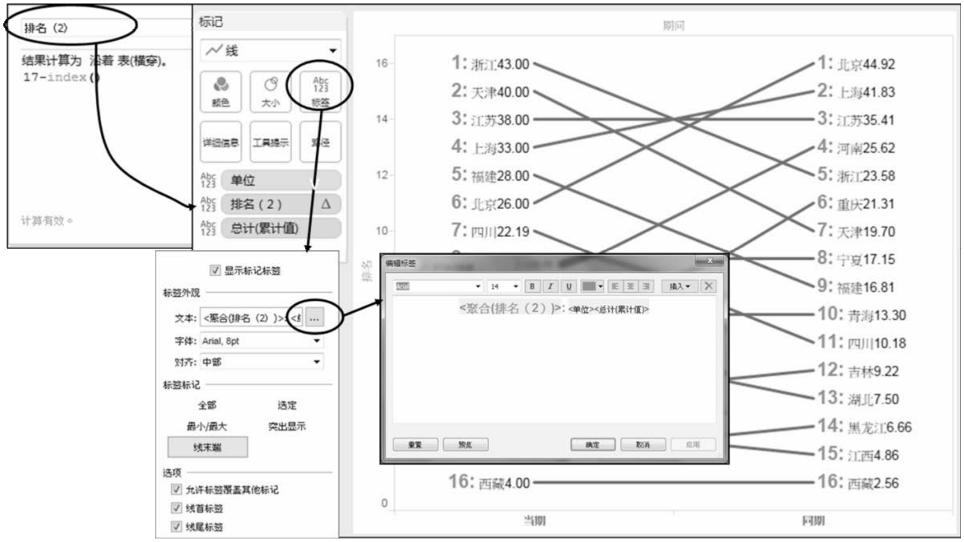


图7-35　添加标签

● **步骤3：设置图形的颜色和形状**

(1) 创建字段“变化量”来表示当期值与同期值的差异，计算公式为`LOOKUP(SUM([累计值]),LAST()) - LOOKUP(SUM([累计值]),FIRST())`。

(2) 创建字段“增加/减少”以区分“变化量”的正负，计算公式为`IF [变化量]>0 then '增加' ELSEIF [变化量]<0 then '减小' ELSE '不变' END`。

(3) 创建字段“变化量绝对值”来表示 “变化量”的绝对值，计算公式为`ABS([变化量])`。

(4) 拖放“增加/减少”到“标记”卡中的“颜色”，并将“减少”选为红色，将“增加”选为绿色；拖放“变化量绝对值”到“大小”；取消纵轴的“显示标题”，结果如图7-36所示。

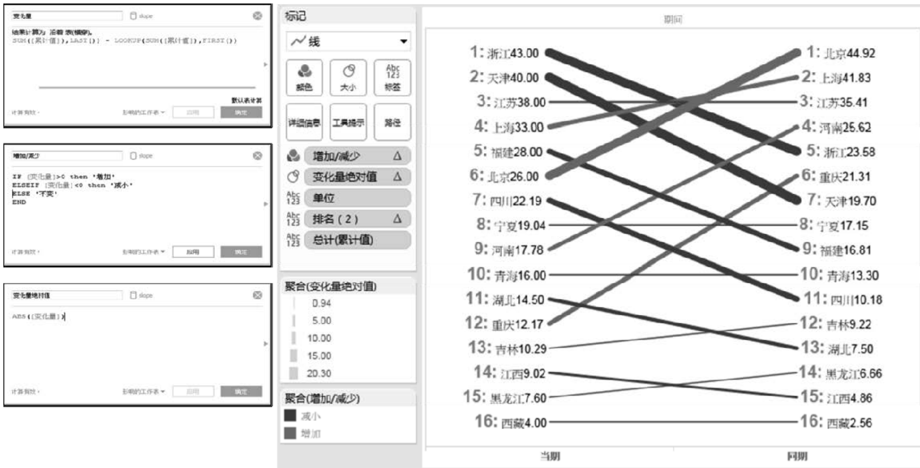


图7-36　设置图形的颜色和形状

7.6　网络图

在分析时，为了更清楚各运行设备的负载情况，需要基于GIS信息展示各变电站、线路的实际地理位置及其负载信息。如图7-37所示，我们用网络拓扑图展示主变压器及输电线路的地理位置，用颜色表示负载的大小；当鼠标移至变电站时，提示变电站名称、所属单位、主变的数量及各主变的负载情况；当鼠标移至线路时，提示线路名称、使用年数、负载率、负载类别等详细信息。

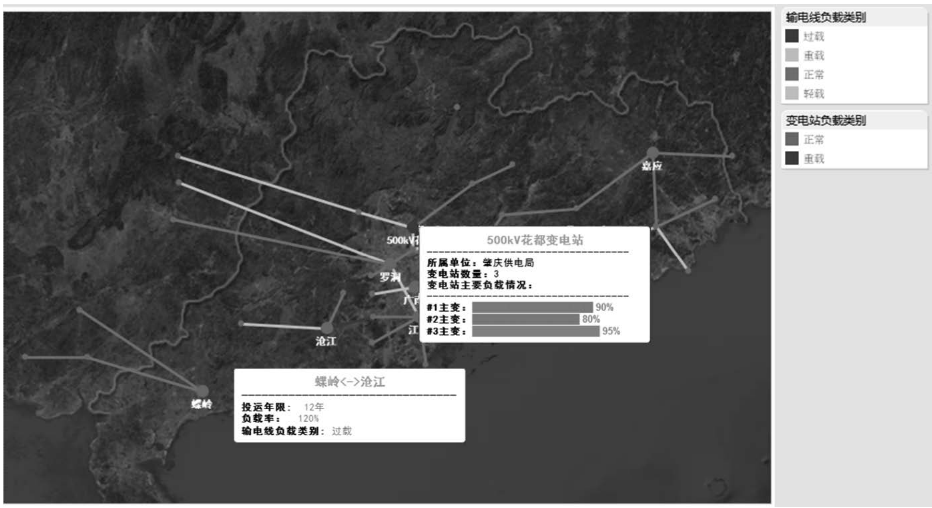


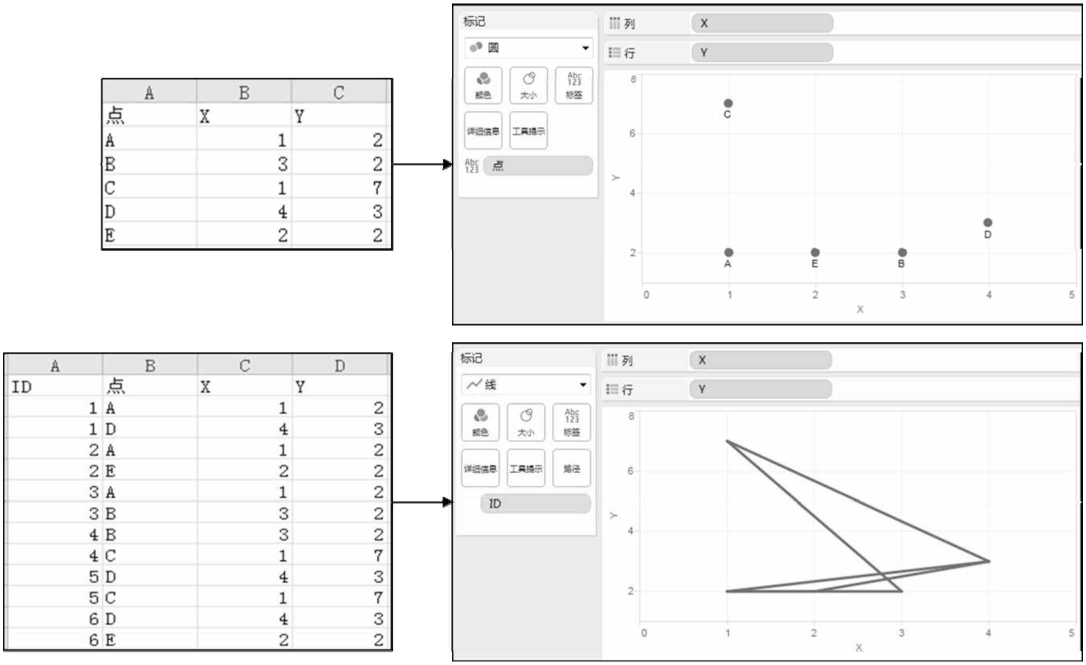
图7-37　设置图形的颜色和形状（另见彩插图7-37）

网络图主要由两个图组成：线图和圆图。线图利用指定点连线的原理做出输电线路的各条路线，圆图则做出主变电站的位置。在创建复杂网络图之前，我们首先介绍线图的连接原理。

1. 指定点连线原理介绍

网络图的难点在于如何让两点进行连线，下面以5个点的简单案例介绍如何实现连接效果。

以图7-38中所示数据结构为例，X、Y为对应的横纵坐标，已知5点的位置可生成包含5个点的点图。如要实现B和C点的连接，则需告知Tableau连接的依据，因此新增字段ID，对B、C两点设置其ID均为4，这样Tableau会将ID相同的点进行连线，即实现了B、C两点的连接。同理，设置其他待连接点的ID，最终生成图7-38（左下）所示的新数据，最终实现AD、AE、AB、BC、DC和DE的连接，如图7-38（右下）所示。



A	B	C
点	X	Y
A	1	2
B	3	2
C	1	7
D	4	3
E	2	2

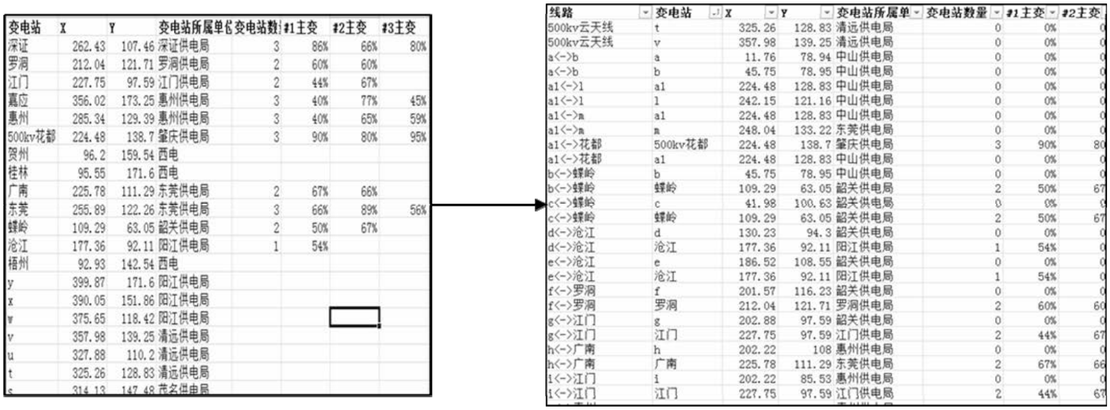
A	B	C	D
ID	点	X	Y
1	A	1	2
1	D	4	3
2	A	1	2
2	E	2	2
3	A	1	2
3	B	3	2
4	B	3	2
4	C	1	7
5	D	4	3
5	C	1	7
6	D	4	3
6	E	2	2

图7-38　指定点连线原理

2. 数据准备

本节将基于变电站数据介绍图7-37所示网络图的创建方式。变电站数据中原始字段包括变电站名称、变电站所属单位、变电站数量、#1主变、#2主变、#3主变、负载率和投运年限。

本案例的数据准备工作较为繁重，首先需要在导入背景图片后依据5.3.2节介绍的方法生成各变电站的坐标X、Y，如图7-39（左）所示。然后我们根据“指定点连原理”确定连接依据 ID字段，在本例中为“线路”字段，其业务含义为连接两个变电站的输电线路名称，如变电站t和变电站v由“500kv云天线”连接，最终数据结构如图7-39（右）所示。



变电站	X	Y	变电站所属单位	变电站数	#1主变	#2主变	#3主变
深证	262.43	107.46	深证供电局	3	86%	66%	80%
罗洞	212.04	121.71	罗洞供电局	2	60%	60%	
江门	227.75	97.59	江门供电局	2	44%	67%	
嘉应	356.02	173.25	惠州供电局	3	40%	77%	45%
惠州	285.34	129.39	惠州供电局	3	40%	65%	59%
500kv花都	224.48	138.7	肇庆供电局	3	90%	80%	95%
贺州	96.2	159.54	西电				
桂林	95.55	171.6	西电				
广南	225.78	111.29	东莞供电局	2	67%	66%	
东莞	255.89	122.26	东莞供电局	3	66%	89%	56%
蝶岭	109.29	63.05	韶关供电局	2	50%	67%	
沧江	177.36	92.11	阳江供电局	1	54%		
梧州	92.93	142.54	西电				
y	399.87	171.6	阳江供电局				
x	390.05	151.86	阳江供电局				
w	375.65	118.42	阳江供电局				
v	357.98	139.25	清远供电局				
u	327.88	110.2	清远供电局				
t	325.26	128.83	清远供电局				
s	314.13	147.48	茂名供电局				

线路	变电站	X	Y	变电站所属单	变电站数量	#1主变	#2主变
500kv云天线	t	325.26	128.83	清远供电局	0	0%	0
500kv云天线	v	357.98	139.25	清远供电局	0	0%	0
a<->b	a	11.76	78.94	中山供电局	0	0%	0
a<->b	b	45.75	78.95	中山供电局	0	0%	0
a1<->l	a1	224.48	128.83	中山供电局	0	0%	0
a1<->l	l	242.15	121.16	中山供电局	0	0%	0
a1<->m	a1	224.48	128.83	中山供电局	0	0%	0
a1<->m	m	248.04	133.22	东莞供电局	0	0%	0
a1<->花都	500kv花都	224.48	138.7	肇庆供电局	3	90%	80
a1<->花都	a1	224.48	128.83	中山供电局	0	0%	0
b<->蝶岭	b	45.75	78.95	中山供电局	0	0%	0
b<->蝶岭	蝶岭	109.29	63.05	韶关供电局	2	50%	67
c<->蝶岭	c	41.98	100.63	韶关供电局	0	0%	0
c<->蝶岭	蝶岭	109.29	63.05	韶关供电局	2	50%	67
d<->沧江	d	130.23	94.3	韶关供电局	0	0%	0
d<->沧江	沧江	177.36	92.11	阳江供电局	1	54%	0
e<->沧江	e	186.52	108.55	韶关供电局	0	0%	0
e<->沧江	沧江	177.36	92.11	阳江供电局	1	54%	0
f<->罗洞	f	201.57	116.23	韶关供电局	0	0%	0
f<->罗洞	罗洞	212.04	121.71	罗洞供电局	2	60%	60
g<->江门	g	202.88	97.59	韶关供电局	0	0%	0
g<->江门	江门	227.75	97.59	江门供电局	2	44%	67
h<->广南	h	202.22	108	惠州供电局	0	0%	0
h<->广南	广南	225.78	111.29	东莞供电局	2	67%	66
i<->江门	i	202.22	85.53	惠州供电局	0	0%	0
i<->江门	江门	227.75	97.59	江门供电局	2	44%	67

图7-39　数据准备

3. 创建线图

● **步骤1：添加背景图片**

依据5.3.2节中所介绍的方法，我们在Tableau中添加网络图的背景图片（参见图7-40）。

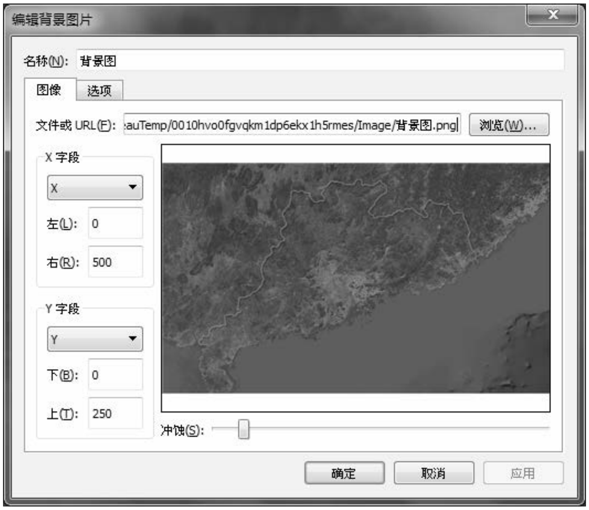


图7-40　添加背景图片

● **步骤2：生成基本线图**

(1) 拖放“X”到列功能区，“Y”到行功能区，并且将其显示方式改为“维度”。

(2)“标记”卡处的标记类型改为“线”图，拖放“线路”字段到“标记”卡处，如图7-41所示。



图7-41 添加背景图片

● **步骤3：改变颜色，添加工具提示**

(1) 创建计算字段“输电线负载类别”，将线路的负载率分为“轻载”“正常”“重载”“过载”4种类型。创建好后拖放到“颜色”，并编辑合理的颜色。

(2) 拖放“投运年限”和“负载率”到“标记”卡中，作为工具提示的内容。因为改变数据结构后，每条线路有两行数据，因此将聚合计算的方式改为“平均值”。

(3) 单击“工具提示”，对格式和内容进行设置。结果如图7-42所示。

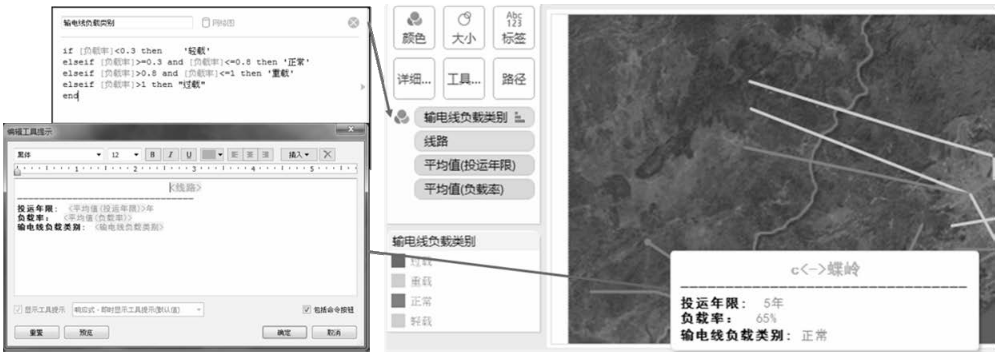


图7-42 添加背景图片

4. 创建圆图

● **步骤1：生成基本圆图**

(1) 在图7-42基础上，拖放字段“Y”到行，并且将第二个视图的标记类型改为“圆”，拖放变电站到“标记”卡处，这时每个点代表一个变电站。

(2) 源数据中变电站较多，而本节只想显示蝶岭、沧江、广南等大的变电站，因此创建“大小”字段，为每个变电站赋予不同的“值”，以实现突出显示部分变电站的效果。在实际应用中，我们可将变电站的运行容量作为判断变电站大小的依据。将“大小”拖放到“标记”卡中，将聚

合计算方式改为“平均值”，(参见图7-43)。

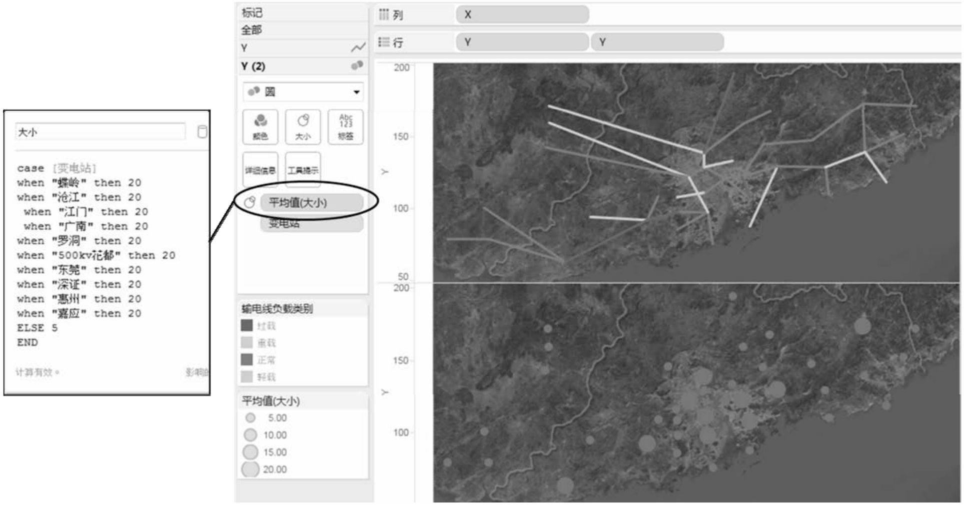


图7-43　创建基本原图

● **步骤2：添加圆图颜色**

变电站一般有多个主变压器，其中任意一个主变的负载率大于80%即为重载。因此，我们创建“变电站负载类别”字段，创建好后拖放到“颜色”，并编辑选择合适的颜色，如图7-44所示。

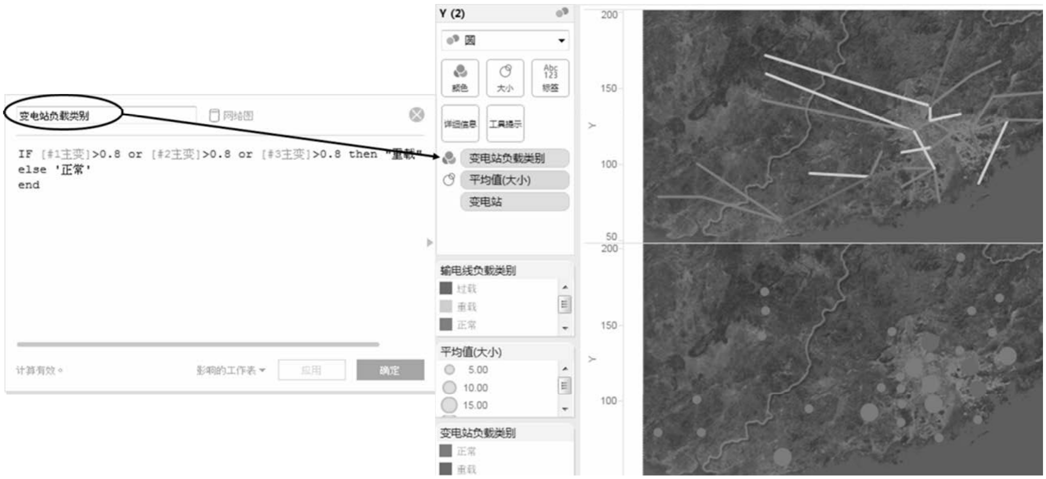


图7-44　添加原图颜色

● **步骤3：添加圆图工具提示**

圆图的工具提示包括变电站名称、所属单位，并且采用条形图直观展示各个主变压器的负载率。

(1) 创建辅助字段。为在工具提示中实现条形图效果，以#1主变为例，先自定义计算字段“#1

百分比”，因为一个变电站在源数据表中存在多个行，对于字段“#1主变”非空的值进行平均运算。随后利用“#1百分比”定义字段“#1柱形图”（计算公式见图7-45），生成条形图字段。同理对#2主变、#3主变创建同样的字段。

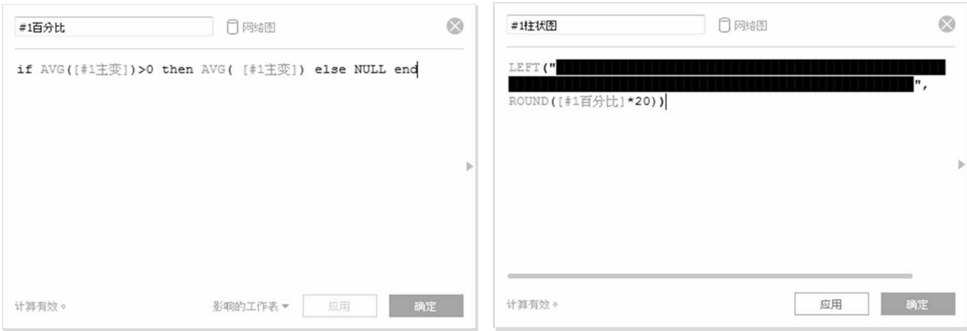


图7-45　创建工具提示字段

(2) 拖放字段“#1百分比”“#1柱形图”“#2百分比”“#2柱形图”“#3百分比”“#3柱形图”“变电站所属单位”“变电站数量”到“标记”卡中，单击“工具提示”，对工具提示的内容和格式进行编辑（参见图7-46）。



图7-46　生成工具提示

- **步骤4：双轴合并**

单击行功能区的*Y*字段进行双轴合并；添加变电站标签，隐藏横纵轴标题，形成最终视图（参见图7-37）。

7.7　雷达图

雷达图能够表达多个指标或多个单位的对比情况，比如在图7-47中，表示不同区县在不同供电区域类别中的10个指标的情况。但因在Tableau中绘制雷达图需要准备大量的底层数据，因此建议尽可能少在Tableau中使用雷达图。

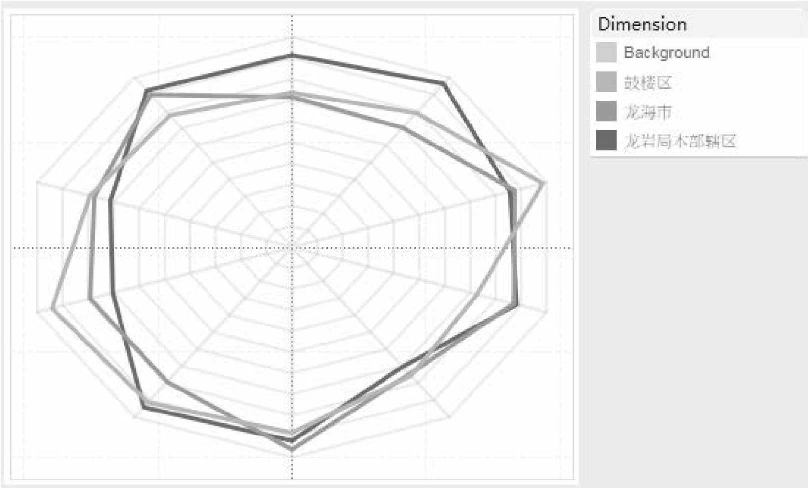


图7-47　雷达图

在Tableau中，雷达图由背景辐射图和业务指标图组成，其中背景辐射图利用指定点连线的原理做出同心圆和辐射线段，业务指标图绘制特定指标的数值。

1. 数据准备

本案例的数据文件包含两部分，一部分是业务指标数据，另一部分则是绘制背景图的数据。

图7-48所示为背景图数据部分，其中Ring代表了同心圆，Radial代表每一个同心圆下所要显示的辐射线段，Spindle Point表示了辐射线段的两个端点。在数据中，Ring为0~10，表示有10个同心圆， Radial为1~21，表示每一个同心圆最多可以有21个辐射线段。

绘制同心圆

Dimension	Metric	Metric Value	统计周期	供电区域	Ring	Radial	Spindle Point
					1	1	
					1	2	
					1	3	
					1	4	
					1	5	
					1	6	
					1	7	

绘制辐射线段

Dimension	Metric	Metric Value	统计周期	供电区域	Ring	Radial	Spindle Point
					0	1	0
					0	2	0
					0	3	0
					0	4	0
					0	5	0
					0	6	0
					0	7	0
					0	8	0
					0	9	0
					0	10	0
					0	11	0

图7-48　背景图数据

表7-1所示为业务指标数据部分，其中Dimension是指标维度， Metric和Metric Value是指标名称和指标值，“统计周期”和“供电区域”是各指标的分析维度。

表7-1 业务指标数据

Dimension	Metric	Metric Value	统计周期	供电区域	Ring	Radial	Spindie Point
龙海市	10kV线路重过载率	85.70375518	2010年	B			
龙海市	10kV线路重过载率	82.73800979	2010年	C			
龙海市	10kV线路重过载率	76.52588844	2011年	B			
龙海市	10kV线路重过载率	96.00157771	2011年	C			
龙海市	10kV线路重过载率	84.11199219	2012年	B			
龙海市	10kV线路重过载率	84.26944644	2012年	C			
龙海市	10kV线路重过载率	96.30576407	2013年	B			
龙海市	10kV线路重过载率	80.06377494	2013年	C			
龙海市	10千伏公用配电线路N–1通过率	87.51469318	2010年	B			
龙海市	10千伏公用配电线路N–1通过率	98.84197765	2010年	C			
龙海市	10千伏公用配电线路N–1通过率	91.07421417	2011年	B			
龙海市	10千伏公用配电线路N–1通过率	72.60662065	2011年	C			
龙海市	10千伏公用配电线路N–1通过率	91.40731392	2012年	B			

为了方便理解数据，表7-2列出了数据源中的字段及其作图含义。

表7-2 数据中各字段含义

字段名称	包含的值	特殊值的含义	说　明
Dimension	区县名称/Null值	Null值表示背景数据	对比维度
Metric	度量指标/Null值	Null值表示背景数据	指标名称
Metric Value	指标值/Null值	Null值表示背景数据	每个维度对应的指标值
统计周期	年份/Null值	Null值表示背景数据	每个维度指标值的年份
供电区域	A+、A、B、C、D及Null值	Null值表示背景数据	每个维度指标值的供电区域
Ring	0~10及Null值	①Null值表示业务数据 ②0表示同心圆圆心 ③1~10表示从内到外第几个同心圆	绘制背景中的同心圆
Radial	1~21及Null值	①Null值表示业务数据 ②1~21表示第几条辐射线段	绘制背景中的同心圆及辐射线段
Spindle Point	0~1及Null值	①Null值表示业务数据及同心圆环上的点 ②0表示同心圆中心点/辐射线端点 ③1表示辐射线段另一个端点	绘制背景中的同心圆及辐射线段

2. 创建雷达背景图

- **步骤1：创建筛选同心圆个数的参数及计算字段**

(1) 创建参数“Number of Rings”，数据类型为“整数”，允许的值为“范围”，“最小值”为1，“最大值”为10，“步长”为1。

(2) 创建用于筛选同心圆个数的字段“Only show enough background rings”，计算公式为`[Ring]<=[Number of Rings]`，即只显示Ring值不大于参数值的环作为背景中的同心圆。

- **步骤2：计算点距离圆心的线段长度**

如图7-49所示，因为同心圆上的点距离圆心的长度相等，且同心圆半径为等差数列，所以第Ring个同心圆上的点距离圆心的长度相当于`([Number of Rings] + 1)-[Ring]`的公差（注意，Ring是从外向内计数），射线长度计算公式见图7-49（右）。



图7-49　线段长度

- **步骤3：计算辐射背景（同心圆上与射线上）点的角度**

(1) 创建字段“Number of Metrics”，计算公式为`SIZE()-1`，结算结果为沿着“Metric”，见图7-50。



图7-50　计算指标最大个数

(2) 创建计算字段“Radials Required”，计算公式为`WINDOW_MAX([Number of Metrics])`，计算结果为沿着“Dimension”“Metric”，见图7-51。

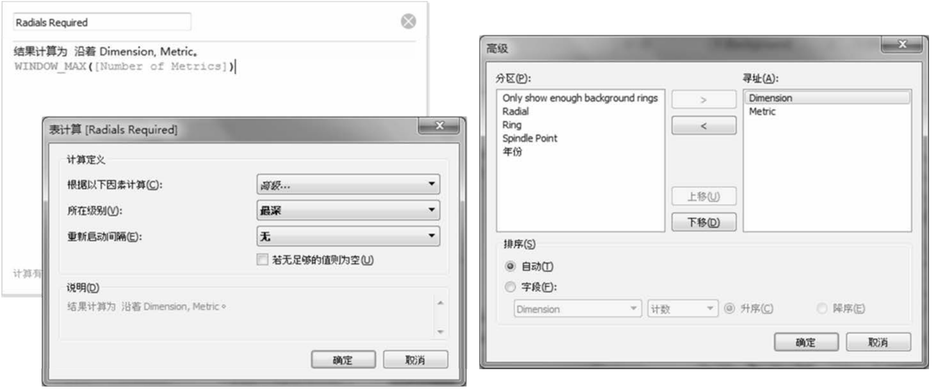


图7-51 需显示指标数

(3) 使用“Radials Required”创建角度计算字段“Angle per point”和“Angle for point (Background)”，计算公式如图7-52所示。

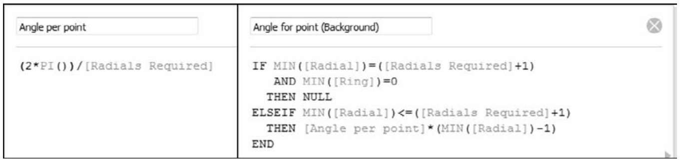


图7-52 角度

这样每个点都有一个“length of line”以及“angle for point(background)”，即得到了每一个点的极坐标。

● **步骤4：利用极坐标计算直角坐标**

创建计算字段“Y(Background)”，计算公式为ATTR([Length of Lines])*COS([Angle for point (Background)])，得到背景图中的*Y*坐标（见图7-53）。同理，创建字段“X (Background)”，计算公式为ATTR([Length of Lines])*SIN([Angle for point (Background)])。



图7-53 创建背景图*Y*坐标

● **步骤5：绘制背景图**

(1) 将“Spindle Point”与“Radial”合并，生成合并字段“Spindle Point & Radial（已合并）”，作为确定背景线图顺序的字段。

(2) 将“Dimension”“Metric”“Ring”以及“Spindle Point & Radial（已合并）”拖放到“标记”卡上。

(3) 将“X”与“Y(Background)”拖放到横列功能区，设置“Y(Background)”的寻址字段为“Dimension”“Metric”“Radial”“Ring”和“Spindle Point”。

(4) 将视图转换成“线”图，并使用“Spindle Point & Radial（已合并）”作为线图的“路径”。其中，“Spindle Point & Radial（已合并）”要按照 (0,1)、(1,1)、(0,2)、(1,2)、(0,3)、(1,3)这样的顺序进行排列，如图7-54所示。

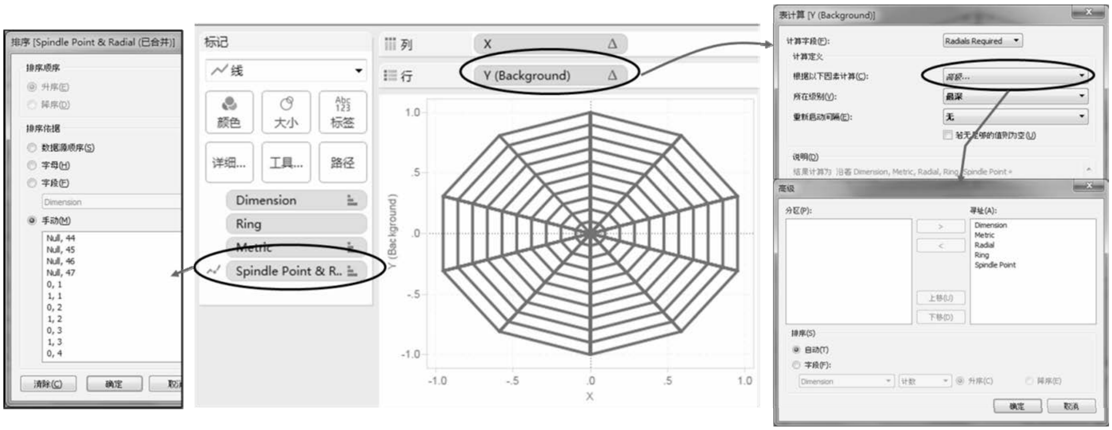


图7-54　创建背景图

3. 创建业务指标图

● **步骤1：计算业务数据中各个指标值的角度**

(1) 创建计算字段“Metric Radial”，结果计算为沿着“Metric”，通过ISNULL()来筛选出“Radial”字段为空的行，即业务指标部分的数据，见图7-55。



图7-55　选中业务指标部分数据

(2) 创建计算字段“Angle for point(Metrics)”，利用“Angle per point”和“Radials Required”计算指标值的角度，如图7-56所示。

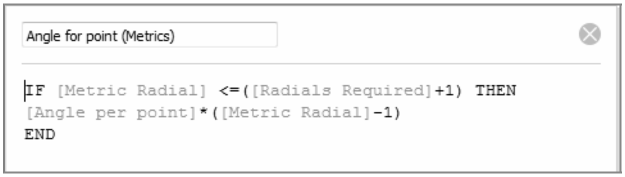


图7-56 业务数据各指标值角度

- **步骤2：利用极坐标计算业务数据指标值的笛卡儿坐标**

(1) 创建计算字段“Max Value”，将每一个“Dimension”下的“Metric Value”归一至[0,1]的单位圆内，结果计算为沿着“Dimension”（参见图7-57）。



图7-57 Max Value

(2) 创建计算字段“Y（Metrics）”，通过余弦将极坐标转化为笛卡儿直角坐标系下的点（参见图7-58）。

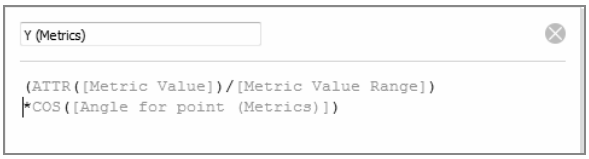


图7-58 笛卡儿坐标

- **步骤3：在背景图上绘制各个维度的指标数值**

将“Y(Mtric)”拖放到行功能区，修改“标记”卡中图形为“线”，“Metric Radial”为“路径”，这样就生成了基本的雷达图，如图7-59所示。可筛选“Dimension”，最终生成如图7-47所示的效果。

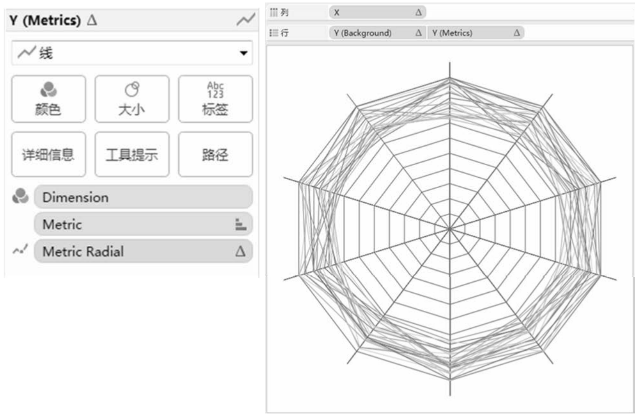

Y (Metrics) Δ
线
颜色
大小
标签
详细信息
工具提示
路径
Dimension
Metric
Metric Radial
列
X
行
Y (Background)
Y (Metrics)

图7-59　生成雷达图

第 8 章

统计分析

前面讨论了如何利用Tableau的基本特性和高级特性创建各种视图，本章主要介绍Tableau自身及集成统计分析与数据挖掘工具R语言之后在预测建模分析方面的能力。

8.1节介绍如何创建各种散点图，以及如何运用散点图分析变量（度量字段）之间的相关关系。

8.2节介绍回归分析及相关模型的基本概念，以及如何在散点图上拟合回归线，还详细介绍了如何进行回归模型的评价。

8.3节介绍如何利用Tableau自带的指数平滑方法针对时间序列数据进行预测，并以坐席接听统计数据中人工服务接听量的预测结果为例介绍了预测模型的评价方法。

R语言是专用于探索和理解数据、预测分析和数据可视化的语言。8.4节介绍了R语言的基本知识，以及在Tableau中集成R语言的详细方法和具体步骤，并以坐席细分为示例介绍如何在Tableau中集成R语言进行主成分分析和坐席人员细分。

本章所用到的数据为“坐席接听统计数据”，其中每条记录包括一个坐席在每天平均接听电话情况的统计信息，以及该坐席的所属中心、部门、组、班、工号、姓名等基本信息。

8.1 散点图与相关分析

散点图是一种常用的表现两个连续变量或多个连续变量之间相关关系的的可视化展现方式，通常在相关性分析之前使用。借由散点图，我们可以大致看出变量之间的相关关系类型和相关强度，理解变量之间的关系。

8.1.1 创建基本散点图

在Tableau中创建基本散点图，需在行列功能区上都放置一个度量。以坐席接听统计数据为例，为直观显示“人工服务接听量”与“服务评价满意率”两个连续变量之间的关系，创建散点图的步骤如下。

(1) 把所需度量字段“人工服务接听量”与“服务评价满意率”分别拖至行功能区和列功能区，此时视图区会把这两个度量按照 “总计”聚合，如图8-1（左）所示。



图8-1　生成散点图

(2) 选择菜单“分析”➤“聚合度量”，移除选中标记，即解聚这两个度量字段，视图区将会以散点图形式显示数据源中的所有数据（每一个标记是数据源中的一行记录）。

8.1.2　创建高级散点图

上一节通过简单操作创建了一个基本的散点图，如果想了解每个坐席人员平均呼入通话时长与人工平均服务接听量之间的关系，可以通过创建合并字段和计算字段来实现。

(1) 创建计算字段“平均呼入通话时长”，计算公式为SUM([呼入通话时长(秒)])/SUM([人工服务接听量])，参见图8-2。之后把 “平均呼入通话时长”拖至列功能区，把“人工服务接听量”拖至行功能区并按平均值聚合。



图8-2　创建计算字段“平均呼入通话时长”

(2) 在“维度”窗口，按住Ctrl同时选中“中心”和“部”，在右键弹出的菜单上选择“合并字段”，创建合并字段“部&中心（已合并）”，右键单击选择“编辑合并的字段”，在弹出的对话框中编辑合并字段的名称为“中心与部”。

(3) 把合并字段“中心与部”拖至“标记”卡上的“颜色”，并把“工号”字段拖放至“标记”卡上，单击 “大小”调节视图的显示效果（参见图8-3）。

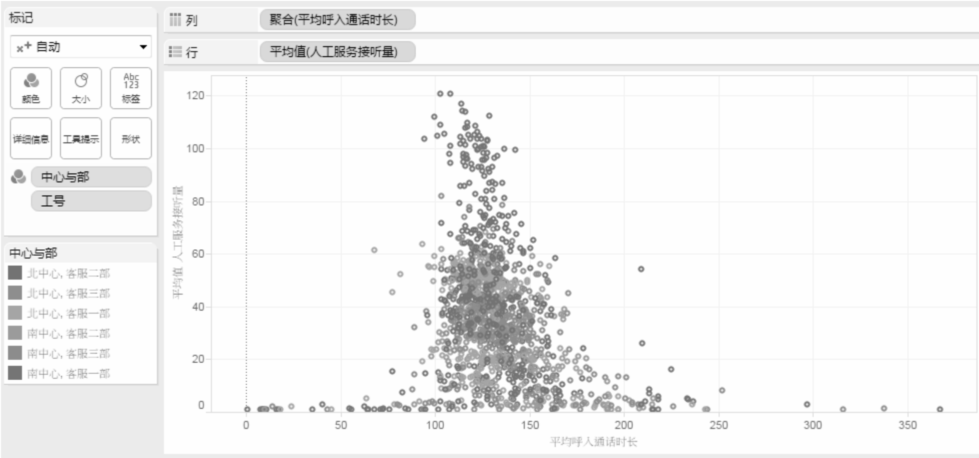


图8-3　创建高级散点图（解聚后）

(4) 单击横轴，在弹出的“添加参考线”对话框中选择依据“平均呼入通话时长”的平均值添加横轴参考线，同时为了使参考线更加醒目，把参考线的颜色由默认颜色改为红色（参见图8-4）。

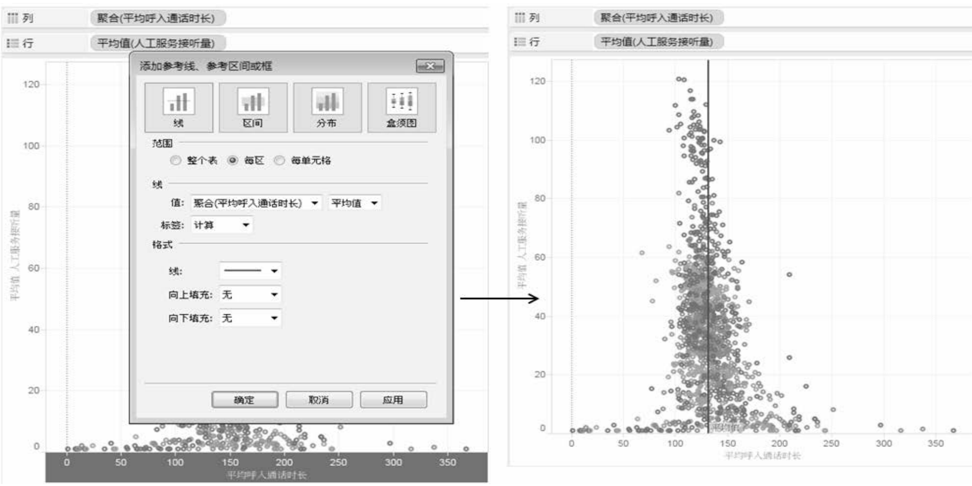


图8-4　为散点图增加参考线

从图8-4可看出：①平均呼入通话时长服从正态分布，均值为131.9 s，绝大多数坐席人员的平均通话时长集中在100 s~160 s；②平均人工服务接听量的数据分布比较分散。

8.1.3　创建散点图矩阵

散点图矩阵是散点图的高维扩展，可以帮助探索两个及以上变量的两两关系，从一定程度上

克服了在平面上展示高维数据的困难，在数据探索阶段具有非常重要的意义。

(1) 把所需度量字段“三声铃响接听量”“不满意数”“人工服务接听量”“平均呼入通话时长（秒）”“平均呼入案头时长（秒）”等分别拖至行功能区和列功能区，并通过菜单“分析”➤“聚合度量”对各个度量进行解聚，参见图8-5。

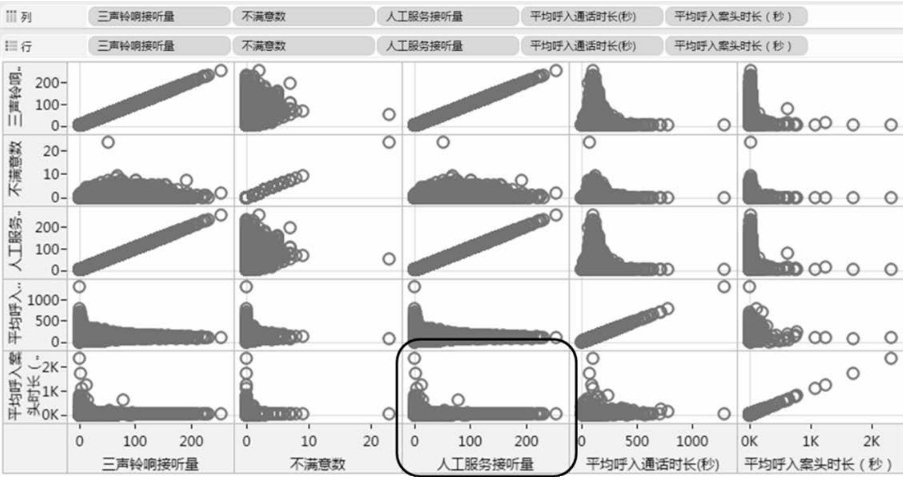


图8-5　创建散点图矩阵（解聚后）

从图8-5可以看出，行功能区和列功能区变量次序相同，对角线上的散点图是一条直线，代表同一变量之间的线性关系，对角线上半部分和下半部分相同。以“平均呼入通话时长”和“人工服务接听量”为例，可以看出“人工服务接听量”数据分散度非常高，而“平均呼入通话时长”分散度很低，随着“人工服务接听量”数据大于一定阈值，“平均呼入通话时长”的集中度更高。

(2) 把合并字段“中心与部”拖至“标记”卡上的“颜色”，即可创建如图8-6所示的散点图矩阵，不同颜色分别代表不同“中心”不同“部”的组合。

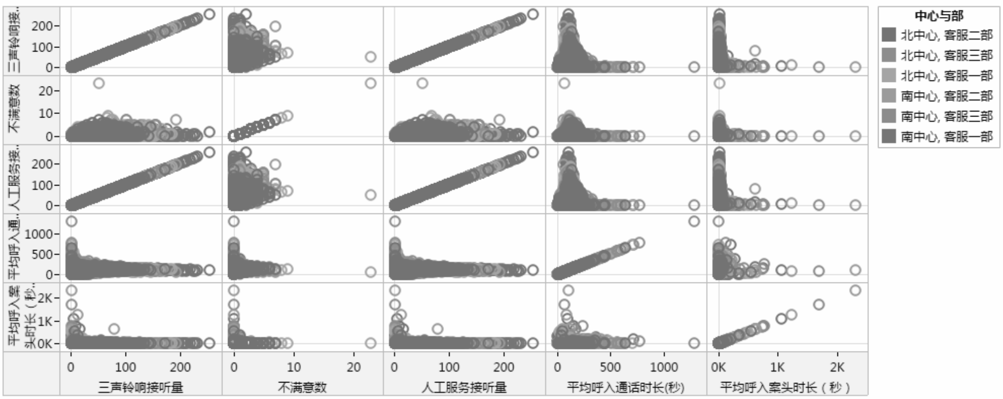


图8-6　利用合并字段调整散点图矩阵

8.2　回归分析

本节利用Tableau的趋势线进行回归分析。在创建散点图之后，通过观察图形，我们可通过添加趋势线对可能存在相关关系的变量进行回归分析，拟合其回归直线或曲线。通过这种分析，我们可以将视图中的趋势线延伸至已有事实数据以外，预测未来值。

8.2.1　模型简介

在向视图添加趋势线时，Tableau将构建一个回归模型，即趋势线模型。通过趋势线模型可以对两个变量的相关性进行分析，通过相关系数及其显著性检验（p值）可以衡量相关关系的密切程度。**显著性检验**指两个变量之间是否真正存在显著的相关关系：只有显著性水平较高时，相关系数才是可信的；相关系数值越大，表示相关性越强。

Tableau中内置了线性模型、对数模型、指数模型和多项式模型等趋势线模型。

- **线性模型**。不对解释变量或自变量执行转换，公式为：$Y = b_0 + b_1 \times X + e$。
- **对数模型**。解释变量或自变量在构建模型之前需进行对数转换，公式为：$Y = b_0 + b_1 \times \text{In}(X) + e$。由于不能对小于零的数字定义对数，因此在估算模型之前解释性变量为负的值将会被筛选掉。因此分析时应避免使用会丢弃某些数据的模型，除非你明确知道筛选掉的数据是无效的。
- **指数模型**。响应变量或因变量在构建模型之前需进行对数转换，公式为：$\text{In}(Y) = b_0 + b_1 \times X + e$。由于不能对小于零的值定义对数，因此在估算模型之前响应变量为负的值将会被筛选掉。
- **多项式模型**。响应变量被转换为解释变量的多项式序列，公式为：$Y = b_0 = b_1 \times X + b_2 \times X^2 + \cdots + e$。Tableau中，我们可以选择介于 2 和 8 之间的“度”（也称为“次”）。较高的多项式度数会放大数据值之间的差异，带来较高的模型复杂度，从而造成过拟合。一般来说，n次多项式模型生成n–1个弯曲的曲线。

说明　(1) 线性回归统计模型中，X是解释变量（或者预测变量），Y是响应变量，b_0表示截距项，b_i表示相关系数，e表示随机误差。随机误差互不关联，与解释变量也不关联，并且具有相等的方差。相关系数的含义是当其对应的解释变量改变一个单位时，其他所引起的Y的改变量。

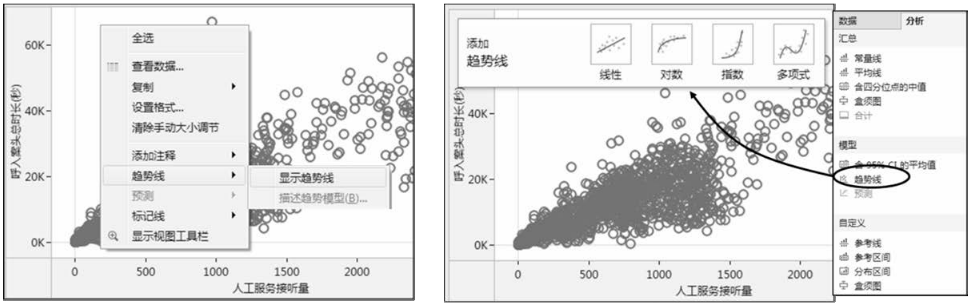
(2) 过拟合是指随着模型复杂度的上升，训练集误差和测试集误差随之下降，但当复杂度到达一临界点后，对训练数据的拟合越来越好，但对于新数据的预测效果逐渐变差。

8.2.2　模型构建

本节将基于实例介绍如何应用Tableau的趋势线功能实现多种回归模型。

(1) 把度量字段“人工服务接听量”“呼入案头总时长（秒）”分别拖至列功能区和行功能区，然后将“工号”拖至“标记”卡中，生成基本散点图。

(2) 为散点图添加趋势线有两种方式：①方式1，在散点图上任意一点单击右键，选择“趋势线”➤“显示趋势线”；②方式2，拖放“分析”窗口处的“趋势线”到视图中。这两种方式的区别是：方式1默认构建线性模型，方式2可选择构建模型的类型（参见图8-7）。



方式1　　　　方式2

图8-7　两种方式创建趋势线

(3) 以“线性模型”为例，生成趋势线后将鼠标悬停在趋势线上，这时可以查看趋势线方程和模型的拟合情况。

如图8-8所示，拟合的线性方程为“呼入案头总时长（秒）= 16.0037 *人工服务接听量+ 1179.95”，显著性p值 < 0.0001。其中1179.95是截距，16.0037是回归系数，含义是“人工服务接听量”每增加一个单位，因变量“呼入案头总时长（秒）”将增加人工服务接听量的16.0037倍。p值是检验模型是否存在显著的相关关系的指标：p值越小，代表模型的显著性越高；一般p值等于或小于0.05代表模型是显著的。

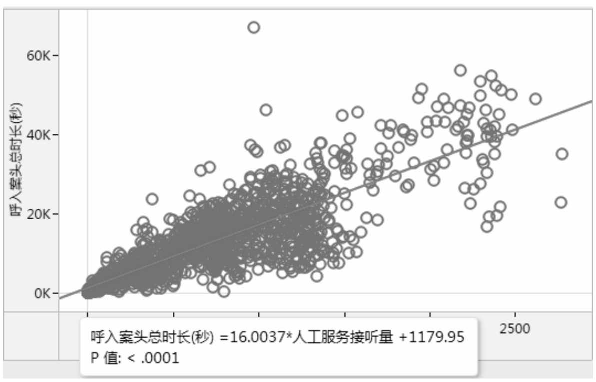


图8-8　线性模型

(4) 在视图上任意一点单击右键，选择“趋势线”➤“编辑趋势线”，Tableau打开“趋势线选项”窗口，此时可编辑趋势线（参见图8-9）。

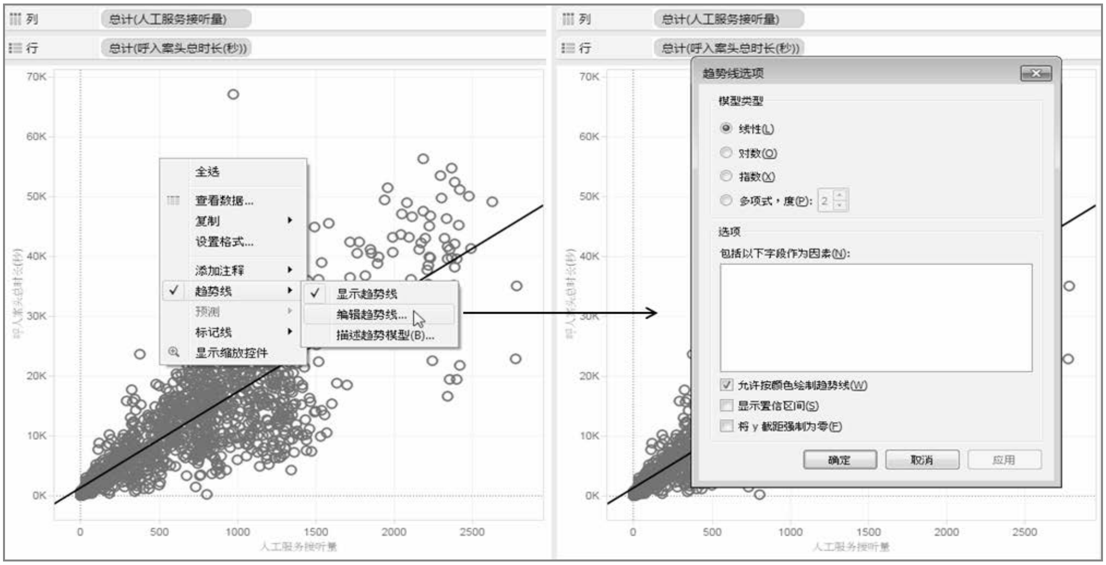
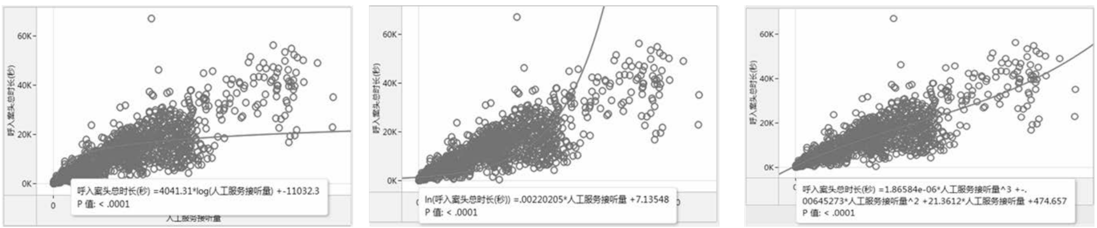


图8-9　编辑趋势线

在“趋势线选项”窗口上，我们可以选择“线性”“对数”“指数”或“多项式”模型类型。“显示置信区间”会显示上和下95%置信区间线，但“指数”模型不支持置信区间。如果需要让趋势线从原点开始，可以设置“将*Y*截距强制为零”。

选择不同的模型类型，生成的趋势线效果见图8-10。



对数模型　　　指数模型　　　多项式模型

图8-10　不同的趋势线模型

(5) 按颜色绘制趋势线。Tableau还允许按照颜色绘制趋势线。首先勾选“趋势线选项”窗口中的“允许按颜色绘制趋势线”选项，然后把“中心”字段拖到“标记”卡中的“颜色”上，即在散点图上增加两条趋势线，其中橙色趋势线拟合南中心员工的数据，而蓝色趋势线拟合北中心员工的数据，参见图8-11。

8

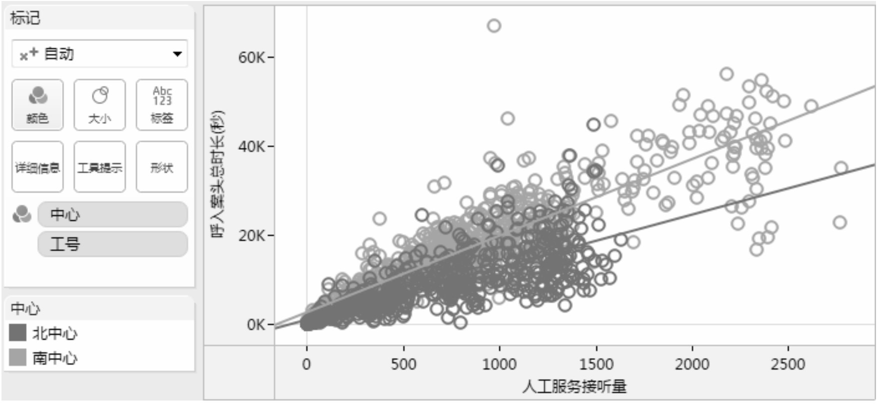


图8-11　按颜色绘制趋势线（另见彩插图8-11）

(6) 有两种方式可以移除已绘制的趋势线：①方式1，通过选择“分析”➤“趋势线”➤“显示趋势线”移除选中标记；②方式2，在视图区中的趋势线上右击，在弹出窗口上移除“显示趋势线”选中标记。通过以上方式移除，下次启用趋势线时将会保留这些趋势线选项。但若在不显示趋势线的情况下关闭工作簿，则趋势线选项会恢复为默认设置。

8.2.3　模型评价

在添加趋势线后，若想查看模型的拟合优度，我们只需在视图中任意一点右击并选择 “趋势线”➤“描述趋势模型”来打开“描述趋势模型”页面，如图8-12所示。

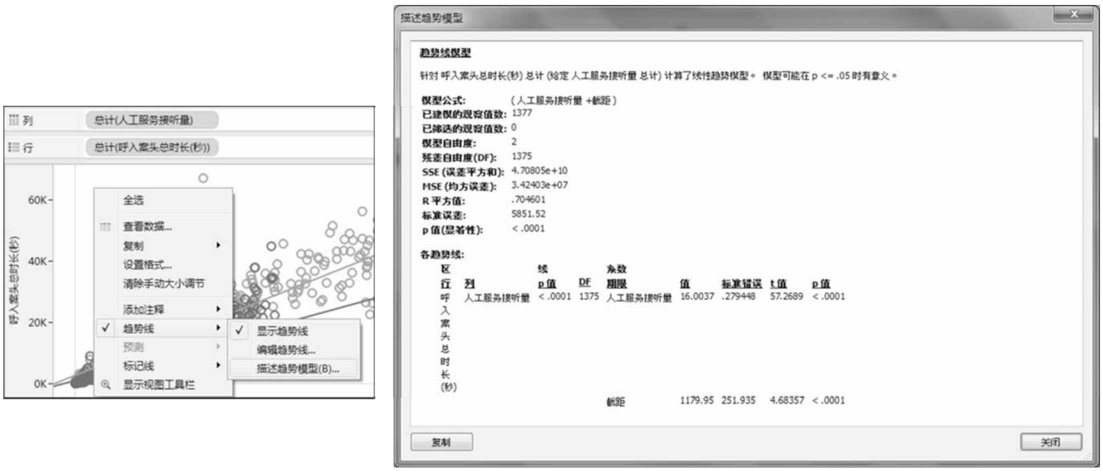


图8-12　趋势线模型评价信息

通过图8-12中的各统计量数据，我们可以获取以下对模型的主要评价信息。

(1) **模型自由度**。指定模型所需的参数个数。线性、对数和指数趋势的模型自由度为2。多项式趋势的模型自由度为 1 加上多项式的次数。

(2) **R平方值**。模型的拟合优度度量，用于评价模型的可靠性，数值大小可以反映趋势线的估计值与对应的实际数据之间的拟合程度，取值范围为0~1。如值为0.704601，则表明模型可以解释服务评价满意率70.4601%的方差。

(3) ***p*值（显著性）**。模型显著性*p*值越小代表模型的显著性越高，值小于0.000 1说明该模型具有统计显著性。*p*值大于0.05可以得出该度量字段与响应变量（服务评价满意率）无关的结论。本例中只用到一个用作预测变量的度量字段，即人工服务接听量，该度量字段*p*值小于0.000 1，表明回归系数显著。

8.3 时间序列分析

Tableau内嵌了对周期性数据的预测功能，可自动拟合预测模型，分析数据变化规律，定量预测数据，同时也可对预测模型的参数进行调整，并评价预测模型的精确度。但Tableau嵌入的预测模型主要考虑数据本身的变化特征，无法充分考虑外部影响，因此适用于周期波动特征明显的数据预测。

8.3.1 时间序列图

时间序列图是一种特殊的散点图，时间作为横轴，纵轴放置不同时间点上变量的取值，可以帮助我们直观地了解数据的变化趋势和季节变化规律。

要绘制时间序列图，我们需要按照时间顺序采集足够多的样本数据，否则它很难反映出数据的规律，时间单位可以是年、季度、月、日，也可以是小时、分钟等。

下面以坐席接听统计数据为例，创建各个中心各客服部一个月之内每一天的人工服务接听总量的时间序列图。

(1) 把“日期”字段拖至列功能区，把“人工服务接听量”和“中心”拖至行功能区。Tableau会自动把日期类型字段，按照年、季、月、日等维度调整，本例中自动按“年”维度调整，为此视图上看到的只有两个标记。

(2) 在列功能区上的“日期”字段上单击右键，在弹出窗口中选择“日”连续日期级别，切换日期字段的级别，之后视图区即显示出1月份31天的时间序列图，如图8-13所示。

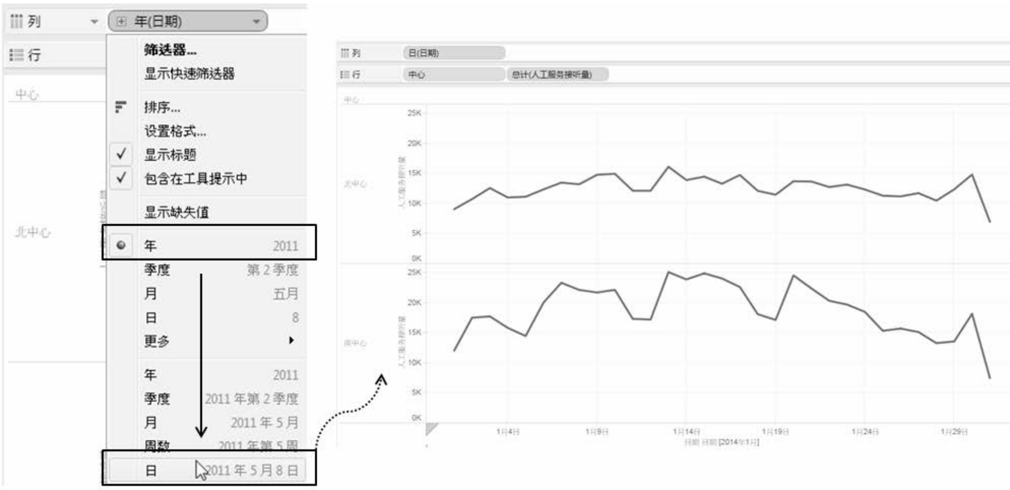


图8-13　切换日期字段的角色和级别

(3) 把“部”拖至“标记”卡中的“颜色”上，视图中的每个区会创建三条时间序列折线，不同颜色代表不同的客服部。可以看出，南中心客服一部每天的人工服务接听量远远高于其他5个客户部，且每天波动较大（参见图8-14）。

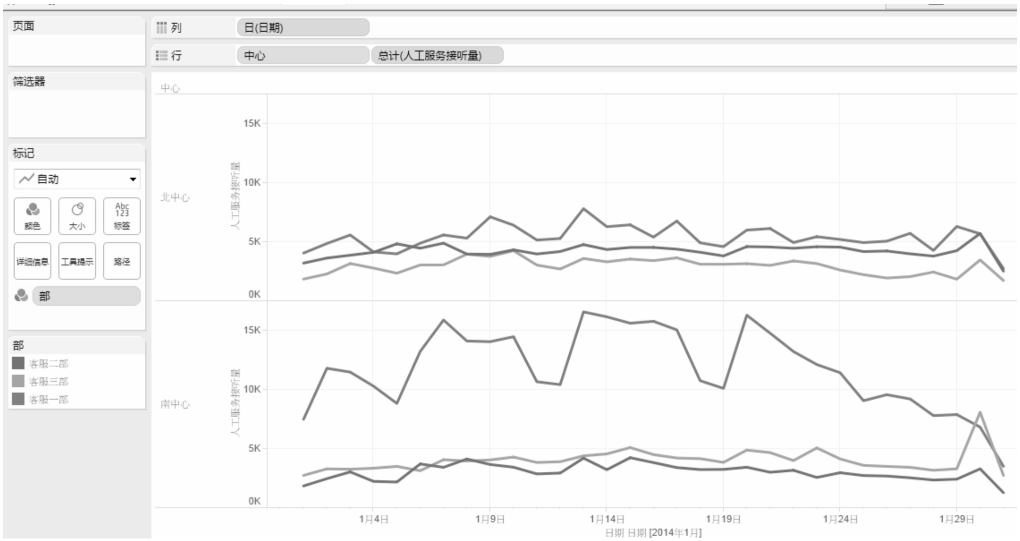


图8-14　为每个中心创建不同客服部的时间序列图

(4) 为了确认人工服务接听量的时间序列数据是否具备“周”波动的特性，为视图增加以周为周期的参考线（为周日添加参考线，一共4条参考线）。

在横轴上单击右键，在弹出窗口上选择“添加参考线”，在“每区”上增加参考线，线的取

值分别为常量值：2014/1/5、2014/1/12、2014/1/19和2014/1/26，如图8-15所示。

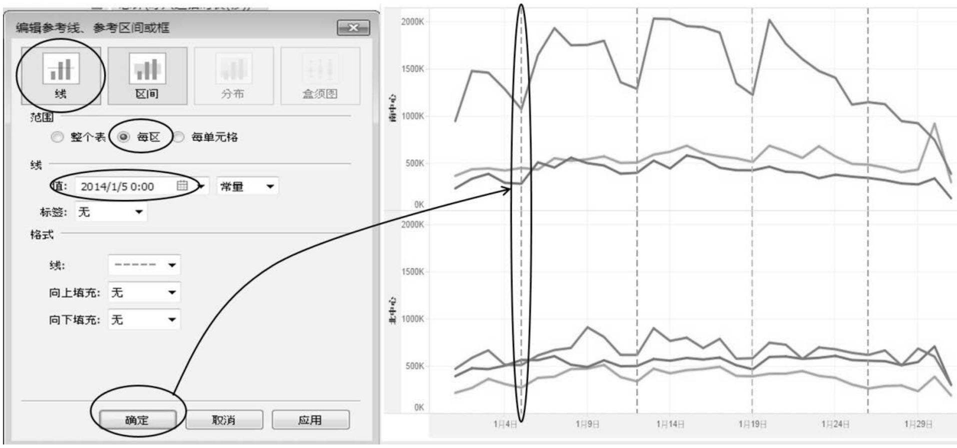


图8-15　添加参考线

由图8-15可知，南中心客服一部以周为周期的变化趋势非常明显，每周的前五天具有稳中上升的趋势，周六周日的接听量急速下降。这可能是因为南中心客服一部服务地区的用户大多数为商业、大工业或非普工业用户。

8.3.2　时间序列预测

时间序列预测是利用原始时间序列数据拟合一个模型，分析研究数据的发展变化规律，从而得出观测数据的统计特征，再依据拟合到的模型外推预测目标的一种定量预测方法。当要预测的度量在进行预测的时间段内呈现出趋势或季节性时，带趋势或季节组件的指数平滑模型十分有效。“趋势”指数据随时间增加或减小的趋势，“季节性”指值的重复和可预测的波动特性。

Tableau嵌入了“指数平滑”的预测模型，即基于历史数据引入一个简化的加权因子，即平滑系数，以迭代的方式预测未来一定周期内的变化趋势。该方法之所以称为指数方法，是因为每个级别的值都受前一个实际值的影响，影响程度呈指数下降，即值越新权重越大。

在预测时，Tableau 针对某个季节周期进行测试，自动找到对估计预测的时间系列而言最典型的时间长度。即如果原始时间序列按月聚合，则 Tableau 将寻找 12 个月的周期；如果原始时间序列按季度聚合，则Tableau 将寻找 4 个季度的周期；如果原始时间序列按天聚合，则 Tableau 将寻找一周的周期。因此，如果按月时间序列中有一个 6 个月的周期，则 Tableau 可能会寻找一个 12 个月模式，其中包含两个类似的子模式。

通常，时间序列中的数据点越多，所产生的预测就越准确。如果要进行季节性建模，那么具有足够的数据尤为重要，因为模型越复杂，就需要越多的数据训练集，这样才能达到合理的精度级别。

下面以坐席接听数据为例介绍创建时间序列预测模型的方法。

(1) 创建“日期”和“人工服务接听量”的时间序列图后即可构建预测模型，Tableau有3种方式生成预测曲线：①方式1为选择“分析”➤“预测”➤“显示预测”；②方式2为在视图上任意一点单击右键，选择“预测”➤“显示预测”；③方式3为拖放“分析”窗口中的“预测”模型到视图中。3种方式创建的效果一致。

预测值在视图中显示在实际历史记录值的右侧，并以其他颜色显示，如图8-16所示。

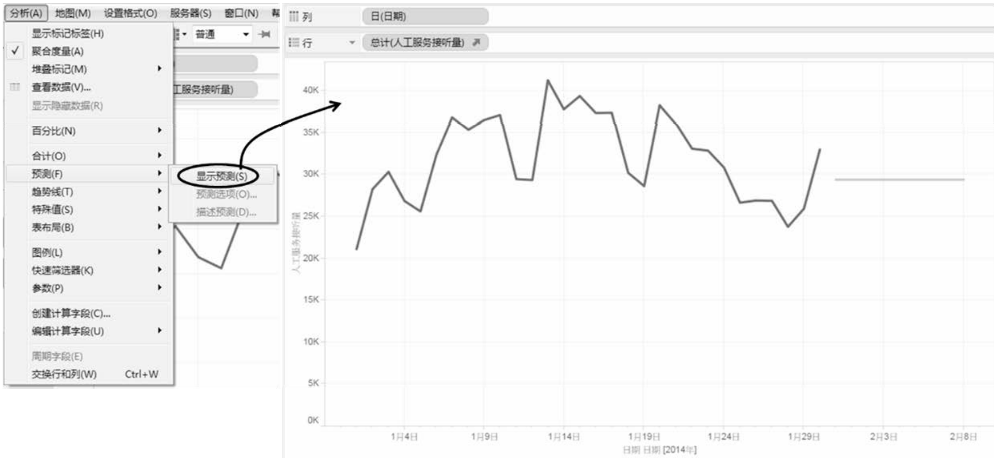


图8-16　方式1创建预测曲线

(2) 改变预测选项，优化预测模型。Tableau默认的预测模型可能不是最优的，如图8-16所示对未来9天的预测值全部是一条直线，不符合期望。Tableau可选择“分析”➤“预测”➤“预测选项”，打开“预测选项”窗口，查看Tableau默认的模型类型和预测选项并进行修改（参见图8-17）。

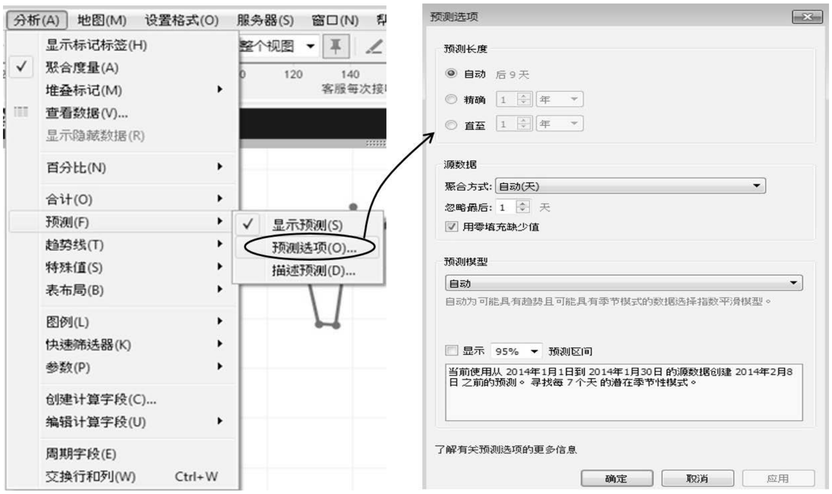


图8-17　修改预测选项

由图8-17可知，预测选项的设置包括以下几项。

1. 预测长度

预测长度功能用于确定预测未来时间的长度，包括“自动”“精确”和“直至”3个选项：①“自动”指自动选取预测长度，本例中Tableau默认选择的是预测未来9天；②“精确”指将预测指定长度和粒度的单位，本例中如果把“精确”选项设置为1月，将会预测2014年2月一个月的人工服务接听值；③“直至”指Tableau将预测扩展至未来的指定时间点，本例中如果把“直至”选项设置为2月，将会预测把人工服务接听量预测到2014年2月28日（包含2月28日）。

2. 源数据

源数据功能用于指定源数据的聚合、期数选取和缺失值处理方式，包括“聚合方式”“忽略最后”和“用零填充缺少值”3个选项。

- “聚合方式”指定时间序列的时间粒度，使用默认值“自动”时，Tableau 将选择最佳粒度进行估算。此粒度通常与可视化项的时间粒度（即预测所采用的日期维度）匹配。当可视化项中的时间序列太短从而无法进行估算时，有时可能并且需要使用比可视化项更精细的粒度来估算预测模型。
- “忽略最后”指定实际数据末尾的周期数，这些周期在估算预测模型时会被忽略。使用这一功能可以修剪掉可能会误导预测的不可靠或部分末端周期。如果经过“忽略最后”选项中指定的周期数筛选后而得到的数据周期数小于5，则在“预测选项”窗口底端会显示“时间序列太短，无法预测”。本例中，实际数据有31期，如果“忽略最后”选项的设置大于等于27，则无法预测。
- “用零填充缺少值”选项指Tableau将会把缺失的期数值填充为0。如果尝试预测的度量缺少值，则 Tableau 将无法创建预测，选择此选项可以防止无法预测的错误发生。

3. 预测模型

预测模型功能用于指定如何生成预测模型，包括“自动”“自动不带季节性”和“自定义”3个选项。

选择“自动”选项，Tableau会选择它认为最佳的模型（本例即是如此）。选择“自动不带季节性”选项，Tableau会使用不带季节性组件的最佳模型。选择“自定义”选项，“预测选项”对话框中将会出现两个子选项“趋势”和“季节”，用于指定模型的趋势和季节特征；两个子选项的取值相同，包括“无”“累加”和“累乘”3个取值，参见图8-18。



图8-18 预测模型选项

如果“趋势”选择“无”，该模型将不针对趋势评估数据；如果选择“累加”模型，指多个独立因素的组合影响是每个因素的孤立影响的总和，可以评估时间序列中的数据以获得累加趋势；如果选择“累乘”模型，指多个独立因素的组合影响是每个因素的孤立影响的乘积，可以评估时间序列中的数据以获得累乘趋势。同理，对季节也是同样的设置效果。

如果一个视图中存在多个时间序列，预测模型选择“自定义”选项，将强制使用同一自定义模型所设置的“趋势”选项值和“季节”选项值。另外，如果要预测的度量的一个或多个值小于或等于零，或者甚至是其中一些数据点相对于其他数据点非常接近于零，则不能使用“累乘”模型。

4. 预测区间

我们可以通过“显示预测区间”设置预测的置信区间为90%、95%、99%，或者输入自定义值，并可设置是否在预测中包含预测区间。本例中，如果将预测区间设置为95%，则将在预测图中显示95%预测区间的阴影区域，表示该预测模型已确定人工服务接听量位于预测周期的阴影区域内的可能性为95%。

5. 预测摘要

“预测选项”对话框底部的文本框提供了当前预测的描述。每次更改上面的任一预测选项时，预测摘要都会更新。如果预测有问题，则该文本框会提供错误消息，该消息可能有助于解决问题。

如果尝试在“预测选项”窗口中把“预测模型”选项设置为“自定义”，并把“趋势”子选项设置为“无”，将“季节”子选项设置为“累加”，同时设置“预测区间”为95%，将会有如图8-19所示的预测结果。

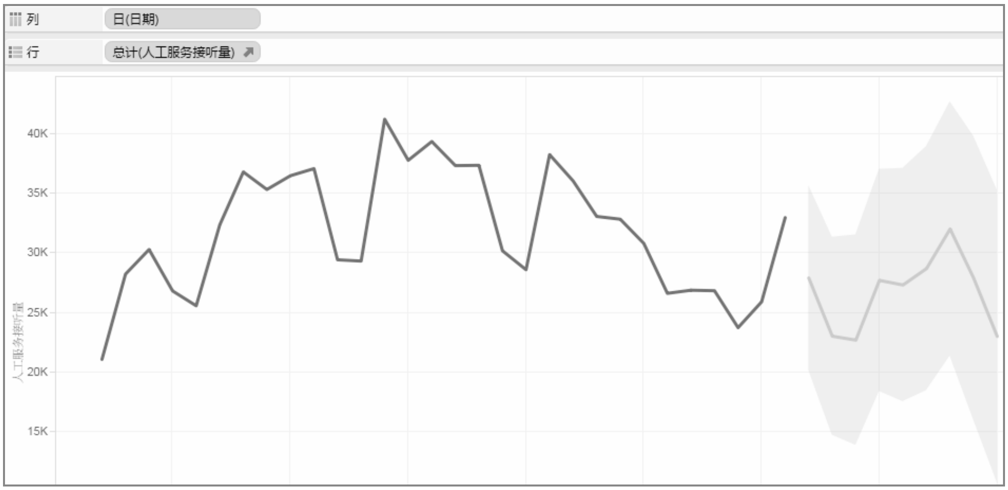


图8-19　优化后的预测结果

说明　Tableau 需要时间序列中具有至少 5 个数据点才能预测趋势，以及具有用于至少两个季节或一个季节加5个周期的足够数据点才能估计季节性。例如，需要至少 9 个数据点才能估计具有一个四季度季节周期（4 + 5）的模型，需要至少 24 个数据点才能估计具有一个 12 个月季节周期（2 * 12）的模型。如果对不具有支持准确预测的足够数据点的视图启用预测，Tableau会为了实现更精细的粒度级别来查询数据源，从而提取足够的数据点来产生有效预测。

8.3.3　预测模型评价

选择“分析”➤“预测”➤“描述预测”打开“描述预测”对话框，这样可以查看关于预测模型的详细描述信息。“描述预测”对话框有两个选项卡：“摘要”选项卡（参见图8-20）和“模型”选项卡。



图8-20　预测模型描述（摘要选项卡）

在“摘要”选项卡中，Tableau描述了已创建的预测模型以及 Tableau 发现的一般模式。上半部分汇总了 Tableau 创建预测所用的选项，一般由 Tableau 自动选取，也可在“预测选项”对话框中指定。

- **时间序列**：用于定义时间序列的连续日期字段。
- **度量**：估计值时使用的度量。
- **向前预测**：预测的长度和日期范围。
- **预测依据**：创建预测所用实际数据的日期范围。
- **忽略最后**：实际数据末尾的周期数将被忽略，该数值用于确定预测数据显示的周期数，此值由“预测选项”对话框中的“忽略最后”选项决定。
- **季节模式**：在数据中找到的季节周期长度。如果在任何预测中都找不到季节周期，则为“无”。

在“摘要”选项卡中，对于预测的每个度量将显示一个摘要表。一般而言，预测摘要表中的字段主要包括以下几项。

- **初始**：第一个预测周期的值和预测间隔。
- **从初始值更改**：第一个和最后一个预测估计点之间的差值（这两个点之间的间隔显示在列标题中）。注意，当值以百分比形式显示时，此字段会显示相对于第一个预测周期的百分比变化。
- **季节影响**：这些字段将针对具有季节性（随时间变化的重复模式）的模型而显示。它们将显示实际值和预测值的合并时间序列中上一个完整季节周期的季节组件的高值和低值。
- **贡献**：趋势和季节性对预测的贡献程度，并且这些值始终以百分比形式表示，且总和为100%。
- **质量**：指示预测与实际数据的相符程度（可能的值为好、确定和差）。

说明 自然预测的定义为：下一周期的值估计将与当前周期的值相同。质量以与自然预测相比较的结果表示。例如，“OK”表示相比自然预测，预测误差更小；“GOOD”表示预测误差要小一半以上；而“POOR”则表示预测的误差更大。

在“模型”选项卡中，Tableau提供了更详尽的统计信息以及霍尔特–温特斯（Holt Winters）指数平滑模型的平滑系数值。对于预测的每个度量将显示一个表，用来描述 Tableau 为该度量创建的预测模型。

“模型”选项卡中包含“模型”“质量指标”以及“平滑系数”，见图8-21。



图8-21 预测模型描述（模型选项卡）

- “模型”指定“级别”“趋势”或“季节”组件是否是用于生成预测模型的一部分，并且每个组件在创建整体预测值时，可以是“无”“累加”与“累乘”。

- “质量指标”中提到的 5 个指标也是在常规的时间序列预测中经常需要的判断指标：RMSE（均方误差）、MAE（平均绝对误差）、MASE（平均绝对标度误差）、MAPE（平均绝对百分比误差）以及常用的AIC（Akaike信息准则）。
- “平滑系数”的拌合参数（包括Alpha级别平滑系数、Beta趋势平滑系数和Gamma季节平滑系数），是根据数据的级别、趋势或季节组件的演变速率对平滑系数进行优化，使得较新数据值的权重大于较早数据值，这样就会将样本内向前一步预测误差最小化。平滑系数越接近1，执行的平滑越少，从而可实现快速组件变化且对最新数据具有较大依赖性；平滑系数越接近0，执行的平滑越多，从而可实现逐渐组件变化且对最新数据具有较小依赖性。

8.4 Tableau 与 R 语言

Tableau通过四大表计算函数嵌入R语言，从而可以借助R语言强大的数据分析挖掘功能实现高级分析的功能拓展。

8.4.1 R语言简介

R语言（简称R）是一种基于对象的开源软件编程语言，也是现今最受欢迎的数据分析和可视化工具之一，由一个庞大且非常活跃的全球性研究型社区维护，并可运行于多种平台之上，包括Windows、UNIX和Mac OS。R提供了一套完全免费的数据分析解决方案，有非常多优秀的特性和能力，主要体现在以下几个方面。

1. 轻松处理各种数据源

R可以识别和处理手工输入数据，亦可轻松地从各种外部数据源导入数据，这些数据源包括文本文件、数据库管理系统、电子表格、统计软件等（如图8-22所示）。

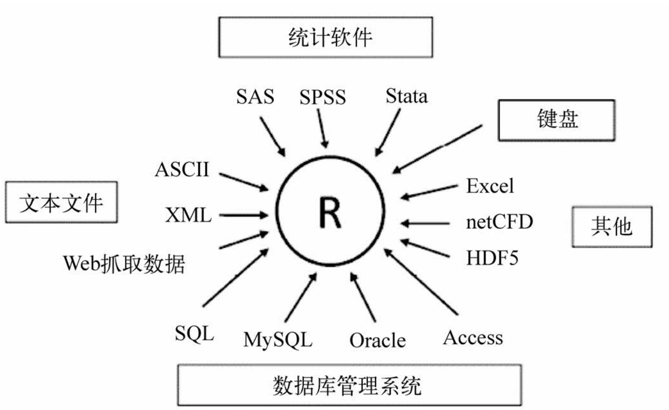


图8-22 可供R导入的数据源

2. 提供丰富的R包

包是R函数、数据、预编译代码以一种定义完善的格式组成的集合。R基本安装中就内置了数以百计的用于数据管理、统计分析和绘图的基础包，如base、datasets、utils、grDevices、graphics、stats及methods等。不过，R更多的增强功能来自R社区开发的数以千计的优秀扩展包，这些包提供了横跨各种领域、数量惊人的新功能，包括地理数据分析和处理、文本数据处理及挖掘、蛋白质质谱分析、心理测验分析等功能。

目前，R的各种扩展包已经超过5000个，可从某个CRAN镜像站点连接http://cran.r-project.org/web/packages以便下载和安装，再在R会话中使用`library()`命令载入下载的扩展包，即可使用扩展包中提供的函数。

3. 交互式的数据分析和探索平台

R是一个可进行交互式数据分析和探索的强大平台，任意一个分析步骤的结果均可被轻松保存、操作，并作为进一步分析的输入，支持典型的数据分析步骤（参见图8-23），几乎可以满足任何类型的数据分析需求。

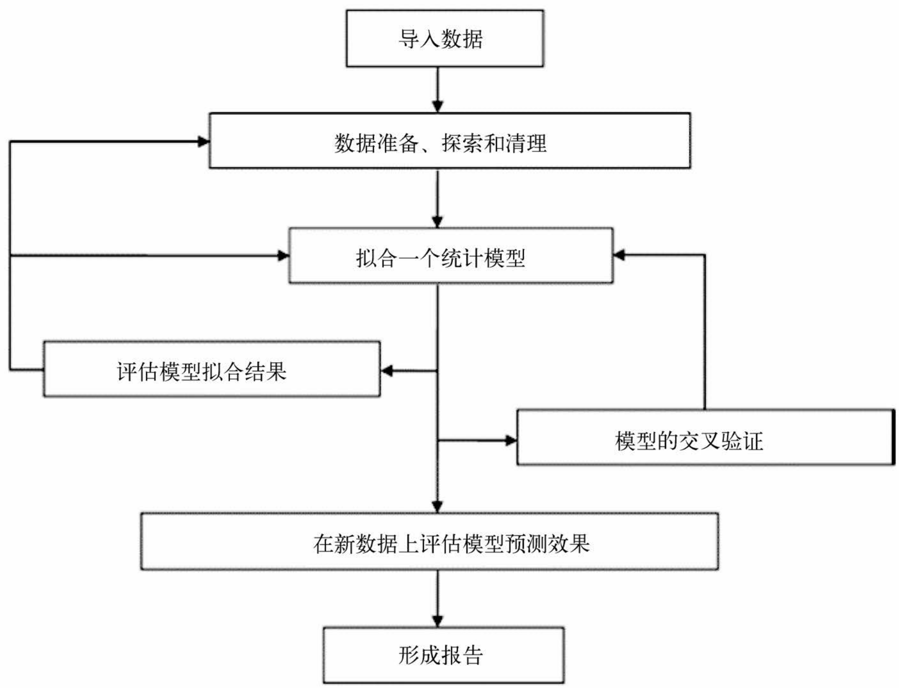


图8-23　典型的数据分析步骤

说明 在R中，对象是可以赋值给变量的任何事物，包括常量、数据结构、函数，甚至是图形。与其他同类软件（如SAS、SPSS和Stata）中的数据集类似，数据框（data frame）是R中用于存储数据的最主要数据结构，其中列表示变量（Tableau中称为字段），行表示观测（记录）。

R的功能非常丰富，但由于许多功能都是由独立贡献者编写的可选模块提供的，相关文档可能比较零散，所以R的学习曲线较为陡峭。另外，要想深入掌握R也需要具有一定统计学知识和一定的编程基础。Tableau完全不受这些方面的限制，可以灵活使用，这是R和Tableau的最重要区别。集成R在数据处理、高级统计分析和预测性建模分析方面的能力可以弥补Tableau在这些方面的不足。

8.4.2　Tableau与R集成

自Tableau 8.1以来，Tableau可以利用四大表计算函数与R的脚本实现集成。具体来说，用户在Tableau环境中，通过Tableau的表计算函数把待处理的数据以参数形式传递给R，R利用函数和R包来处理数据或建立预测模型，然后把处理结果或者预测模型的输出结果通过表计算函数返回给Tableau，最后 Tableau对结果进行可视化展示和分析。Tableau与R的集成可大大增强Tableau在数据处理，尤其是高级统计分析和预测性分析方面的能力。

本节假定用户已经安装好了R，并掌握了R的一些基础知识。要实现Tableau与R的集成，主要需要以下3个步骤。

1. 安装并运行Rserve，启动Rserve服务进程

Rserve基于TCP/IP协议，是允许R语言与其他编程语言通信的程序包，支持C、C++、Java、Python等。所以，如果要实现Tableau与R的集成，我们首先需要使用R中的Rserve程序包。可以在R会话中执行如下代码来获取Rserve程序包：

```
> install.packages(“Rserve”)
```


成功安装Rserve包后就可以加载并运行Rserve，以便启动Rserve服务进程，而启动Rserve服务进程就意味着在计算机上运行了能向应用程序提供R语言功能的服务器（不止是Tableau这样的程序能使用R功能，其他应用程序也可以）。在R会话中键入如下命令即可：

```
> library(Rserve)
> Rserve()
```

第一个语句执行加载Rserve程序包的命令，第二个语句单独启动一个Rserve服务进程，这样就启动了Rserve服务器。用户可以在任务管理器中查看Rserve服务进程是否已启动，也可以手动终止已经启动的Rserve服务进程。

说明　通过使用Rserve，应用程序能借助TCP/IP或本地套接字（Sockets）访问R。其中，在类Unix系统下Rserve服务器将以守护进程(daemon)模式启动，而在Windows系统下，`Rserve()`通过设置环境变量`PATH`包含当前R.dll文件来运行Rserve服务器。
8.4.2节和8.4.3节中所有的R代码均以>开头，且使用加粗斜体标注。

2. 在Tableau中配置 Rserve 连接

为了在Tableau中使用R程序，我们还需要在Tableau中配置Rserve连接，具体步骤如下。

(1) 在 Tableau中的“帮助”菜单上选择“设置和性能”➤“管理 R 连接”以打开连接对话框。

(2) 使用域或 IP 地址输入或选择服务器名称（本机下配置则是localhost），并且指定端口（端口 6311 是 Rserve 服务器的默认端口）。

(3) 如果Rserve服务器需要凭据，请指定用户名和密码。

这样就能成功连接Tableau到Rserve，并进行测试连接了。如果服务器正常，那么就会返回“成功连接到Rserve服务”（如图8-24所示），如果无法建立连接，就会显示错误消息。如果有错误消息，你可以单击消息中的“显示详细信息”查看服务器返回的诊断信息。

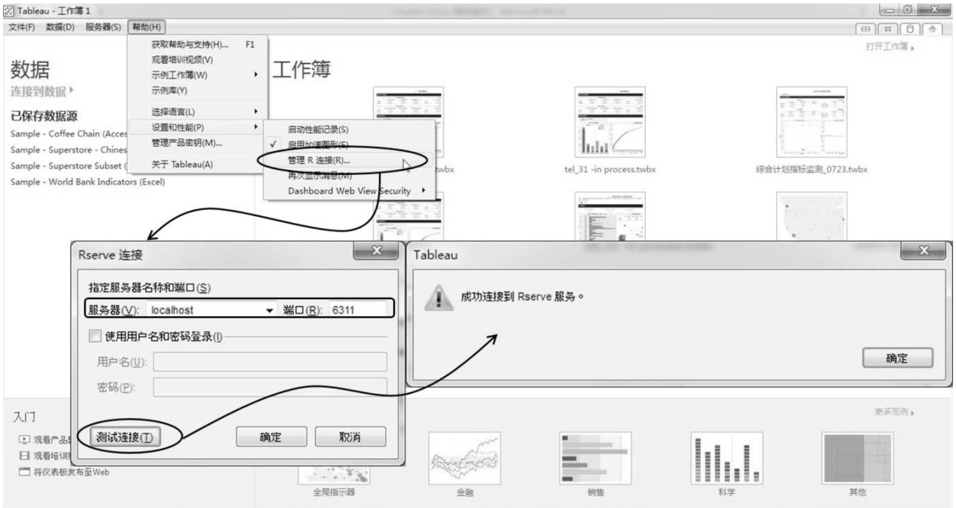


图8-24　将Tableau连接到Rserve

值得注意的是，Tableau 无法验证使用R的工作簿在 Tableau Server 上能否正确呈现。可能存在这样的情形：需要的统计库在用户计算机上可用，但在 Tableau Server 运行的 Rserve 实例上不可用。并且，如果要将依赖 Rserve 连接的工作簿发布到 Tableau Server，应将 Tableau Server 配置为具有自己的 Rserve 连接。我们可通过使用 `tabadmin set` 配置设置来完成此操作（这些设置等同于在“Rserve 连接”对话框中设置的值）：

- tabadmin set vizqlserver.rserve.host;

- tabadmin set vizqlserver.rserve.port；
- tabadmin set vizqlserver.rserve.username；
- tabadmin set vizqlserver.rserve.password；

无论是将包含 R 功能的工作簿发送到其他用户，还是用户自行从 Tableau Server 下载包含 R 功能的工作簿，用户都必须首先在计算机上配置 Rserve 连接，然后才能够在收到或下载到的Tableau工作簿中使用 R 功能。

3. 在Tableau中使用R脚本

在Tableau中，R脚本以表计算函数嵌入到Tableau中，运行时Tableau通过Rserve触发R Engine并传递参数给R，运行结果（数据处理、统计分析和预测建模结果）将以创建新的计算字段的方式返回给Tableau。

Tableau提供了4种R脚本表计算函数：SCRIPT_BOOL、SCRIPT_INT、SCRIPT_REAL、SCRIPT_STR，你可使用这4个函数将 R 脚本传递给 Rserve 服务器并获取结果。

- SCRIPT_BOOL：返回指定 R 表达式的布尔结果。
- SCRIPT_INT：返回指定 R 表达式的整数结果。
- SCRIPT_REAL：返回指定 R 表达式的实数结果。
- SCRIPT_STR：返回指定 R 表达式的字符串结果。

下面以SCRIPT_INT为例介绍如何在Tableau的表计算函数中嵌入R表达式脚本。

语法：SCRIPT_INT('R语言表达式脚本（含参数）', 来自Tableau的传入参数1, 来自Tableau的传入参数2, …)。

在“R语言表达式脚本（含参数）”中，我们编写包含占位符的R语言脚本，可在R语言表达式脚本中通过“.arg*n*”的方式引用来自Tableau的传入参数（来自Tableau的传入参数一般是聚合后的Tableau字段）。同样以SCRIPT_INT为例，我们给出一个进行聚类分析的例子：

```
SCRIPT_INT('result <- kmeans(data.frame(.arg1, .arg2, .arg3), 3);result$cluster;', SUM([A]),
SUM([B]),SUM([C]))
```

对于R语言来说，实际执行的代码如下：

```
>result <- kmeans(data.frame(SUM([A]), SUM([B]),SUM([C])), 3)
>result$cluster
```

熟悉R语言的读者会很清楚，这两行代码首先构造了一个包含3个列向量（.arg1、.arg2、.arg3）的数据框，然后通过K均值算法使用这3个变量对所有观测记录进行聚类，并指定聚类的数量为3类，最后将各观测记录所属聚类结果的类别返回。这里使用了“.arg*n*”的方式来传入参数，将Tableau中字段A、B和C的总计聚合值作为参数，依次替代了R脚本中标示出来的3个占位符.arg1、.arg2和.arg3。

8.4.3 用R进行高级分析

本节仍以坐席接听统计数据为例，介绍如何利用R进行主成分分析和坐席行为细分，具体步骤如下。

1. 分析思路

本节使用的坐席接听统计数据中涉及的字段比较多，包括“工号”“人工服务接听量”“三声铃响接听量”“呼入通话时长（秒）”“呼入案头总时长（秒）”“服务评价推送成功数”和“服务评价满意数”，共计7个字段。其中可能部分属性与数据分析任务不相关，是冗余的。为了更好地依据除“工号”外的6个字段进行聚类分析，我们需要进行数据降维。

具体分析思路：以“工号”为聚合依据，对6个字段使用主成分分析法进行数据降维，最后运用K均值法对坐席人员进行聚类，最后将聚类结果进行可视化展示。

说明　**主成分分析**（principle components analysis）是一种运用线性变换简化数据集的技术，旨在利用降维的思想把多指标转化为少数几个综合指标。为了全面、系统地分析问题，我们必须考虑众多影响因素（变量），因为每个变量都在不同程度上反映了所研究问题的某些信息，并且指标之间彼此有一定的相关性，因而所得的统计数据反映的信息在一定程度上有重叠，所以我们需要主成分分析。

聚类分析（cluster analysis）是根据“物以类聚”的道理，对大量样品（或指标）在没有任何先验知识的情况下进行分类。其中，K均值算法是很典型的基于距离的聚类，采用距离作为相似性的评价指标，即认为两个对象的距离越近，其相似度就越大。

---

2. 借助R进行主成分分析（选取主成分）

进行主成分分析时首先需要分析一些指标来判断如何选取主成分，而这部分内容在Tableau中实现起来比较困难，所以我们使用R来进行主成分分析。执行如下R脚本即可完成主成分分析：

```
> PCA <- prcomp(data, scale = TRUE)
```

确定主成分需要先了解各个主成分的摘要情况（参见表8-1），并依据累计方差累计贡献率达到85%以上（所选主成分对信息的利用率达85%以上）这一准则进行主成分选择。执行如下R脚本可以获得各个主成分的摘要信息：

```
> summary(PCA)
```

表8-1　主成分摘要表

	PC1	PC2	PC3	PC4	PC5	PC6
标准差	2.027	1.038	0.896 2	0.095 23	0.009 561	0.000 443 7
贡献率	0.685	0.179 6	0.133 9	0.001 51	0.000 020	0.000 000 0
累计贡献率	0.685	0.864 6	0.998 5	0.999 98	1.000 000	1.000 000 0

从表8-1主成分摘要表中可以发现，第一主成分PC1和第二主成分PC2的累计贡献率为86.46%，所以我们保留两个PC1和PC2主成分用作聚类分析。

3. 把两个主成分定义为Tableau计算字段

下面为两个主成分创建计算字段，一个命名为“第一主成分”，另一个命名为“第二主成分”

（参见图8-25）。

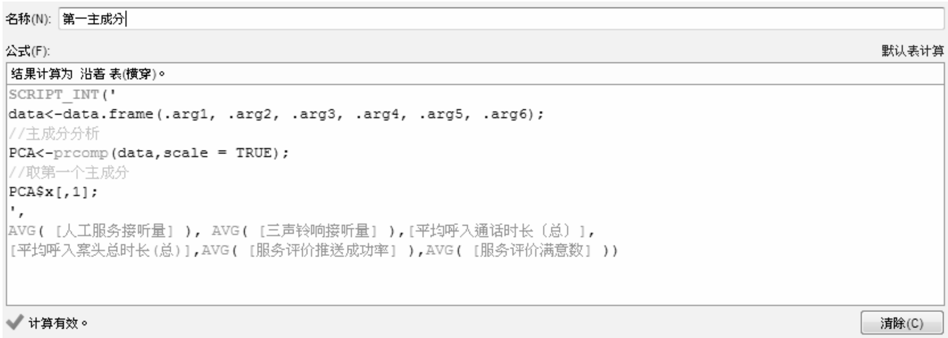


图8-25 创建“第一主成分”计算字段

这里主要使用了R中的语句PCA$x[,1]与PCA$x[,2]来获取主成分分析结果中的第一主成分和第二主成分。第二主成分计算字段的创建与其非常类似，这里不再赘述。

4. 依据两个主成分对坐席人员进行聚类，创建“聚类类别”计算字段

我们采用R中的K均值法，使用两个主成分来将对坐席人员分组，以观察组间行为的差异性及组内的相似性。

首先创建“聚类类别”计算字段，其中的公式如下所示：

```
SCRIPT_INT('result <- kmeans(data.frame(.arg1, .arg2), 4); result$cluster;', [第一主成分], [第二主成分])
```

对坐席人员进行聚类分析的结果如图8-26所示。

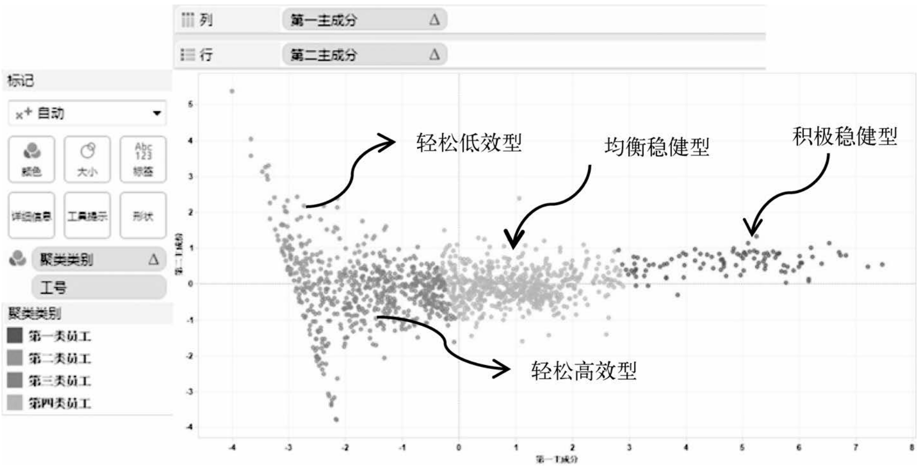


图8-26 坐席人员聚类分析结果

5. 解释成分载荷与聚类结果

为了解释主成分，进而更好地理解聚类结果，我们需要使用R语言进一步分析成分载荷，根据载荷矩阵对这种结果做出某种解释。通过Rsw脚本获得的主成分载荷矩阵如表8-2所示。

表8-2　载荷矩阵

字　段	第一主成分	第二主成分
人工服务接听量	0.99	0.10
三声铃响接听量	0.99	0.10
呼入通话时长	−0.29	0.71
呼入案头总时长	−0.28	0.73
服务评价推送成功数	0.99	0.11
服务评价满意数	0.99	0.11

从表8-2可以看出主成分的含义：第一主成分与人工服务接听量、三声铃响接听量、服务评价推送成功数以及服务评价满意数都是正相关的，而第二主成份与呼入通话时长、呼入案头总时长之间正相关。为了更好地理解，我们可以为这些主成分提供解释，第一主成分代表了坐席人员接听服务的质与量，第二主成分代表了坐席人员处理业务的时间效率。

基于坐席人员行为，我们把坐席人员分为了四类：第一类员工，接听服务质与量高且处理效率正常，代表了“积极稳健型”坐席；第二类员工，接听服务质与量较低且处理效率低，代表了“轻松低效型”坐席；第三类员工，接听服务质与量低但是处理效率较高，代表了“轻松高效型”坐席；第四类员工，接听服务质与量中等且处理效率正常，代表了“均衡稳健型”坐席。

本节着重介绍了如何利用Tableau进行数据分析。我们使用散点图分析变量之间的相关关系，利用趋势线进行趋势分析，运用指数平滑方法建立时间序列预测模型，最后介绍了如何集成R进行主成分分析和聚类分析。

本章介绍的前三节属于Tableau自带的分析功能，虽然简单实用，但是并不能满足复杂的问题分析需求。实际应用过程中，我们需要结合R来完成统计分析和预测建模分析方面的工作。

遗憾的是，Tableau 9.1还不能让用户自由地进行较为高级的探索性分析，主要是由于其开放的R集成接口和集成方式有诸多限制，使得数据处理效率不高。但是我们有理由相信，Tableau逐步发布的更新版本将能够与R完美融合， Tableau将把自由的R带给用户。

第 9 章

分析图表整合

前面探讨了如何利用Tableau进行数据分析、可视化展示，本章主要介绍如何利用Tableau的仪表板功能进行分析图表的整合。

9.1节介绍仪表板相关的一些基本概念，包括对象、布局容器、布局方式、交互操作等。

9.2节通过详细介绍一个实例的设计配置步骤展示仪表板的创建方法。

本章所有分析内容用到的数据源都是“综合计划指标明细表”，每条记录包括指标名称、省市、统计周期和当期值等字段。

9.1 仪表板简介

仪表板指显示在单一面板的多个工作表和支持信息的集合，它便于同时比较和监测各种数据，并可添加筛选器、突出显示、网页链接等操作，实现工作表之间层层下钻、更具交互性的工作成果展示。

9.1.1 工作区

仪表板工作区环境界面如图9-1所示。

- **工作表窗口**。工作表窗口列出当前在工作簿中的工作表，新建工作表后，仪表板窗口会自动更新，这样在添加至仪表板时，所有工作表都始终可用。
- **容器与对象窗口**。对象是指除工作表外可用于辅助监测主题展示的要素，包括图片、文本、网页和空白等。

容器可用于在仪表板中组织工作表和其他对象。新增容器会在仪表板中创建一个区域，在此区域中，对象根据容器中的其他对象自动调整自己的大小和位置。

- **布局窗口**。布局窗口的平铺和浮动选项可用于调整工作表或对象的布局方式，其下方通过树形结构展现了视图中各工作表和对象的层级结构。
- **仪表板窗口**。仪表板窗口可用于调整视图中各工作表或对象等的大小和位置，以及仪表板整体的大小，见图9-2。仪表板默认大小为“台式计算机”，即1000像素×800像素。我们可通过下拉菜单来调整仪表板尺寸，其中“自动”指仪表板自动填充整个窗口，“精确”指仪表板始终保持固定大小，“范围”指仪表板中的所有对象将在指定的最大值和最小值

之间进行缩放并展示。

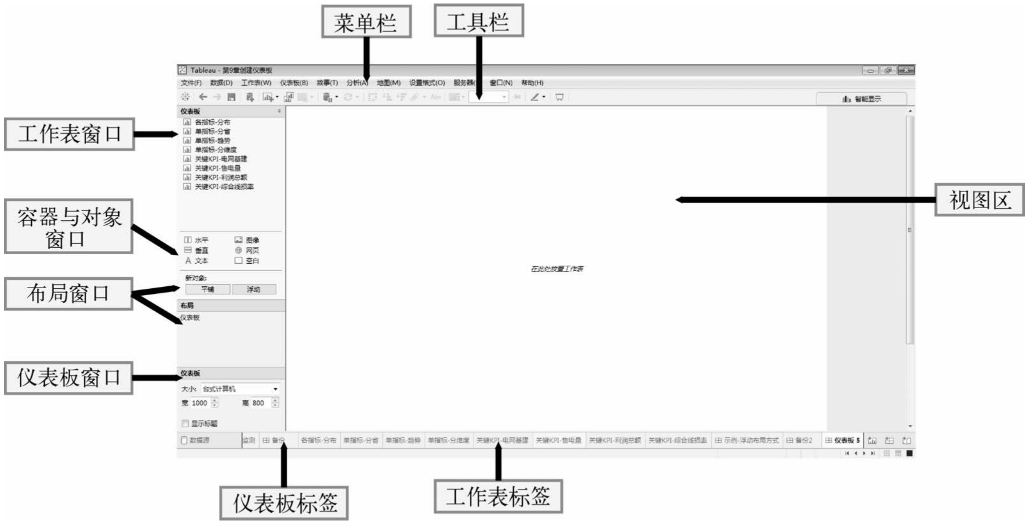


图9-1　仪表板工作区环境



图9-2　调整仪表板大小

- **视图区**。视图区是创建和调整仪表板的工作区域，用户可以向其中添加工作表及各类对象。

9.1.2　对象

在Tableau仪表板中，文本、图像、网页、空白等都可以被当作对象添加至仪表板中，以丰富

展示内容，优化展示效果。

1. 文本

通过文本对象，我们可向仪表板添加文本块，以用于添加标题、说明等。文本对象将自动调整大小，以最佳方式适应仪表板中的放置位置；用户也可以通过拖动文本对象的边缘手动调整其大小。默认情况下，文本对象是透明的，用户可以右击选择设置文本格式。

2. 图像

通过图像对象，我们可向仪表板中添加静态图像文件，如公司Logo或描述性图表。在添加图像对象时，系统会提示从计算机中选择图像，此时可进一步调整图像的显示方式（如大小、对齐方式）并允许为图像添加网页链接。

3. 网页

通过网页对象，我们可将网页嵌入到仪表板中，以便将Tableau内容与其他应用程序中的信息进行组合。添加完成后，链接将自动在仪表板中打开，而不需打开浏览器窗口。

4. 空白

通过空白对象，我们可向仪表板添加空白区域以优化布局，并可通过单击并拖动区域的边缘调整空白对象的大小。

9.1.3　布局容器

布局容器是仪表板的基本构成单元，也是仪表板布局的框架，分为水平和垂直两种，用来放置工作表、快速筛选器、图例、图片、文本、网页等。

1. 水平容器

水平容器为横向左右布局（见图9-3），用户可通过拖放的方式将工作表或对象等添加至其中，添加完成后其宽度会自动调整，以均等填充容器宽度。



图9-3　水平布局容器

2. 垂直容器

垂直容器为纵向上下布局（见图9-4），用户可通过拖放的方式将工作表或对象等添加其中，添加完成后其高度会自动调整，以均等填充容器高度。



图9-4　垂直布局容器

9.1.4　布局方式

布局方式指仪表板中各容器、工作表及图片等其他对象的放置方式，分平铺和浮动两种。

1. 平铺布局

Tableau默认采用的是平铺布局方式（见图9-5），即所选工作表或者对象平行分布而互不覆盖。平铺对象排列在一个单层网格中，Tableau会根据整个仪表板大小和其中的对象自动分布宽度与高度。用户可以通过单击并拖动区域的边缘进行手工调整。

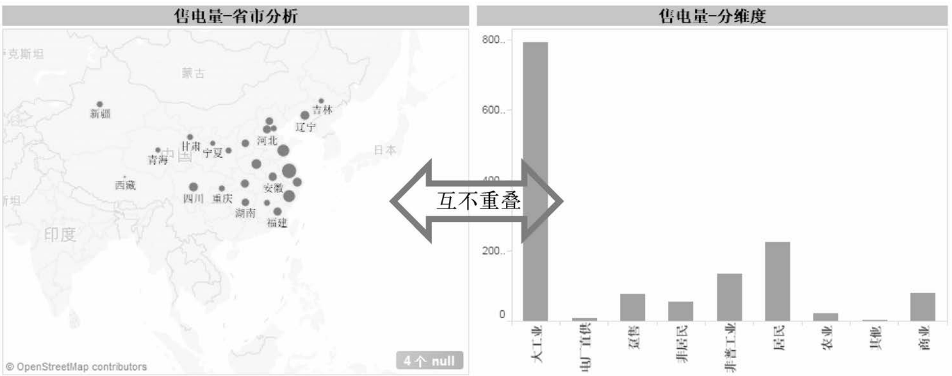


图9-5　平铺布局方式

2. 浮动布局

Tableau还支持采用浮动布局方式（见图9-6），即所选工作表或者对象浮动并覆盖展示于背景视图中，此时用户可选中各对象随意调整其大小与位置。如采用地图等存在较大空白区域的对象时，可以通过采用浮动布局的工作表、快速筛选器、图例、文本、图像等来填充相应的空白区域，以达到更好的展示效果。



图9-6　浮动布局方式

9.1.5　交互操作

交互操作指选中仪表板上的某工作表或对象时，仪表板上的其他工作表或对象也能够与之关联并展示对应的交互内容。Tableau提供了3种灵活的交互方式，包括“表间筛选”“突出显示”“网址链接”，详见9.2.3节。

9.2　操作步骤

本节将详细解析图9-7中实例的配置步骤，借此来介绍仪表板的创建方法。

前提：需提前准备好配置仪表板所需的“零件”——工作表。实例可从图灵社区本书主页http://www.ituring.com.cn/book/1736下载获取，读者也可根据手边数据参照操作。

9

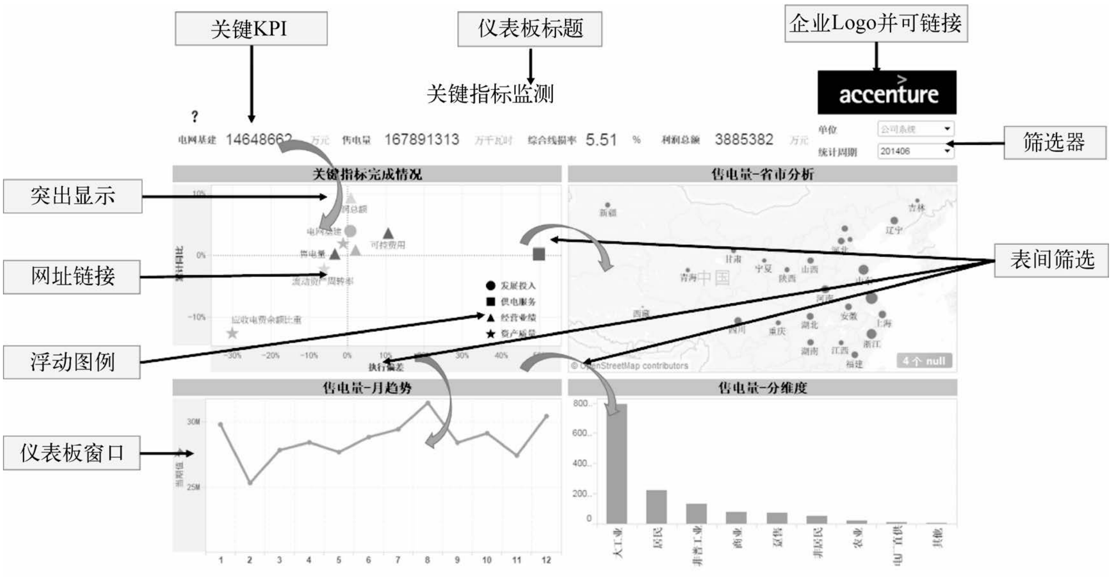


图9-7　仪表板实例

9.2.1　新建布局

新建布局是仪表板配置的首要工作。本节将通过实例介绍如何在Tableau中新建垂直、水平及混合布局方式。

1. 新建仪表板

在菜单栏中选择“仪表板” ➤ “新建仪表板”（见图9-8），或直接单击 “新建仪表板”标签来创建一个新的仪表板。Tableau 将自动生成仪表板名称，第一个仪表板名为“仪表板 1”，第二个为“仪表板2”，以此类推。你可以右键单击仪表板标签并选择“重命名仪表板”，也可以通过双击仪表板标签键入新名称。



图9-8　新建仪表板

2. 分析布局结构

图9-9显示了仪表板实例的布局方式，仪表板实例的区域1、区域3、区域4均为普通的横向水平布局容器，配置时首先添加“水平”容器，再依次添加内容即可。

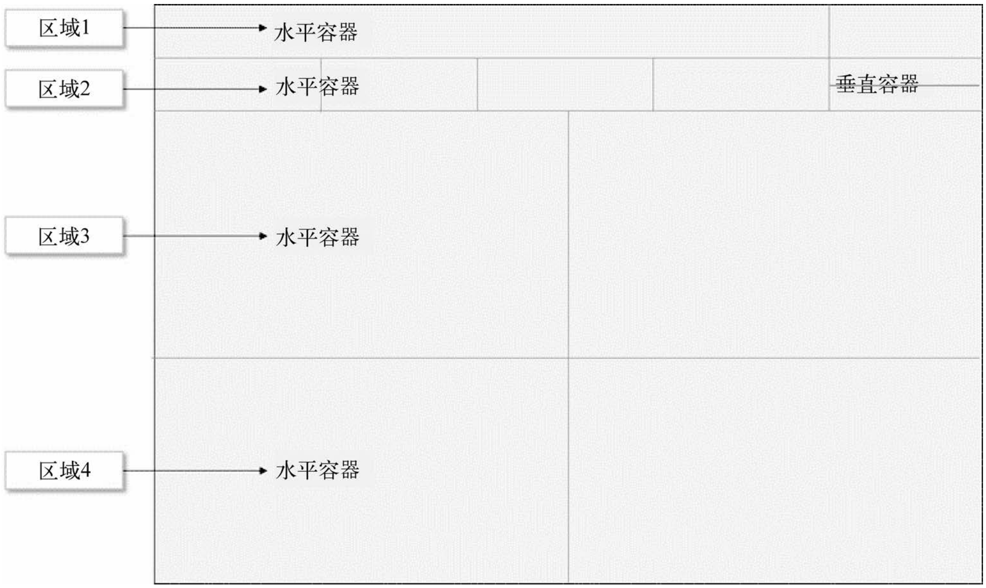


图9-9　仪表板布局方式

图9-9中需特别注意区域2，其布局方式为横向水平布局与纵向垂直布局的组合（见图9-10）。

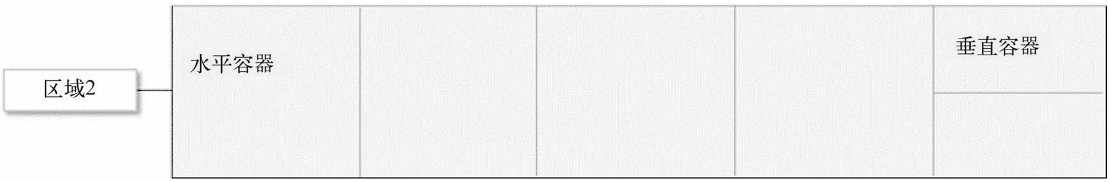


图9-10　区域2的布局方式

在Tableau中布局该区域时必须遵守如下操作顺序：

(1) 添加水平容器；

(2) 在水平容器中添加内容；

(3) 在最右侧添加垂直容器；

(4) 在垂直容器中添加内容。

如果调换(2)、(3)的顺序，仪表板视图区的水平容器会被平铺的垂直容器所覆盖，导致无法添加水平布局方式的内容。以下讲解将按照正确的操作顺序展开，针对可能的错误操作不再做进一步说明。

3. 区域1添加布局容器

从“容器与对象窗口”把“水平”字段拖放至“视图窗口”（见图9-11）。

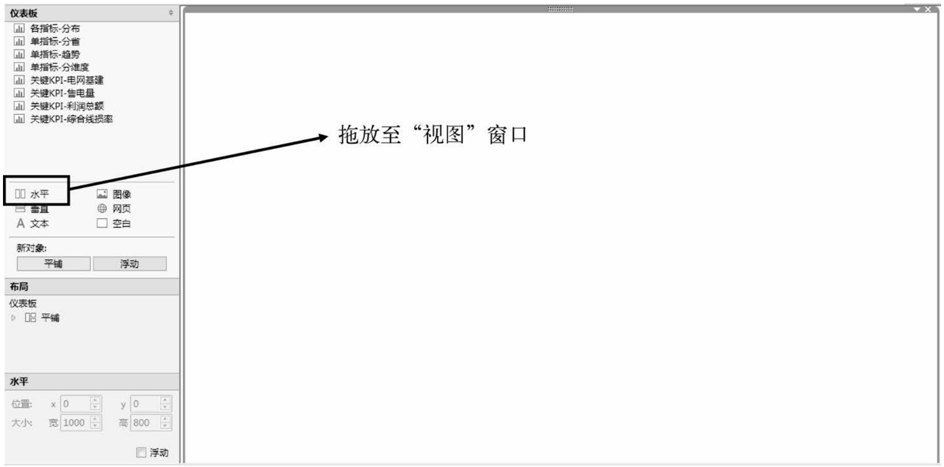


图9-11　添加区域1布局容器

9.2.2　添加内容并调整格式

本节基于实例详细介绍如何向仪表板中添加视图、对象等内容，以及如何进行格式调整以满足分析展示要求。

1. 区域1添加文本生成仪表板标题

从“容器与对象窗口”把“文本”字段拖放至“水平容器”（见图9-12）。

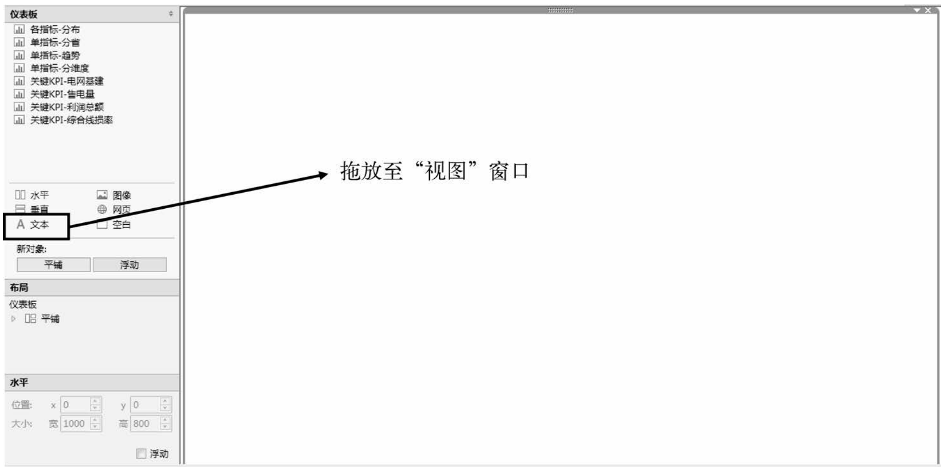


图9-12　区域1添加文本对象

Tableau会自动弹出“编辑文本”窗口，用户可以在其中录入标题名称，同时可以使用与Word类似的工具栏调整字体、字号、粗体、斜体、字体颜色、对齐方式等（见图9-13）。

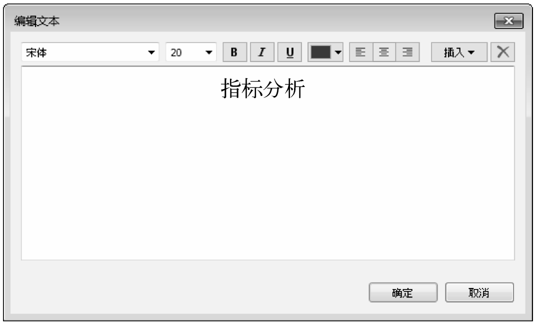


图9-13 在文本对象弹出窗口中录入标题名称

此时由于文本对象的布局方式为平铺，展示效果为充满整个水平容器（见图9-14）。

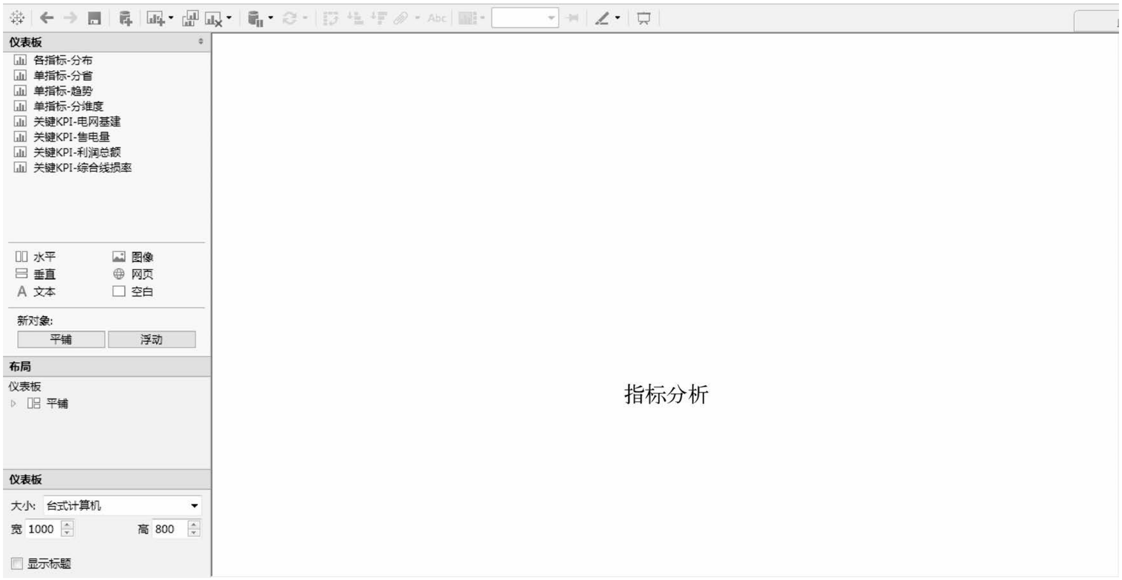


图9-14 录入“标题名称”后的视图窗口

2. 区域1添加公司Logo图像

从“容器与对象窗口”把“图像”字段拖放至“水平容器”中“文本”对象的右侧阴影处（见图9-15）。

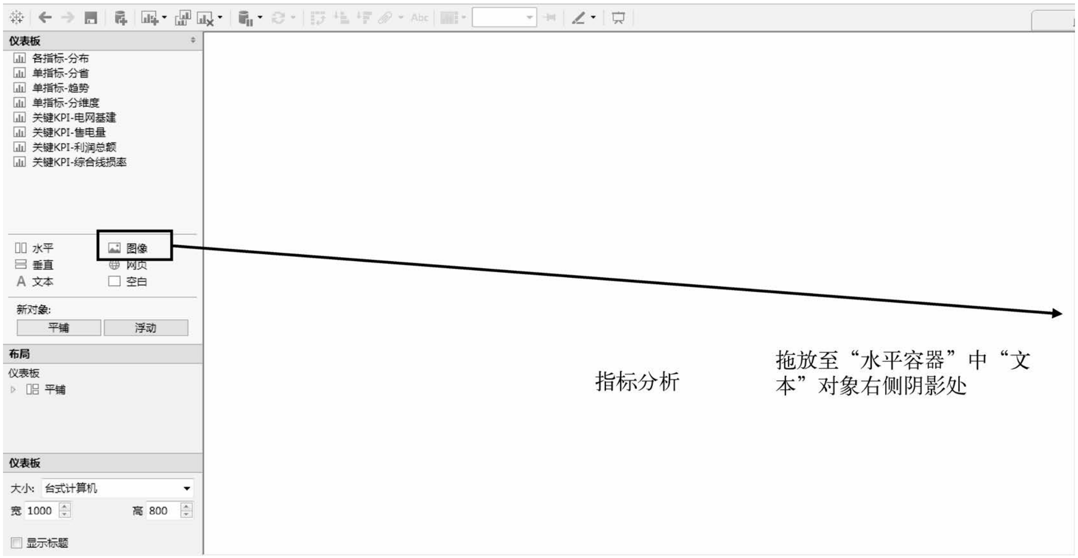


图9-15　添加图像对象

Tableau会自动弹出“图像选择”窗口，用户可以找到存放图像的文件夹位置，选中图片后单击“打开”（见图9-16）。

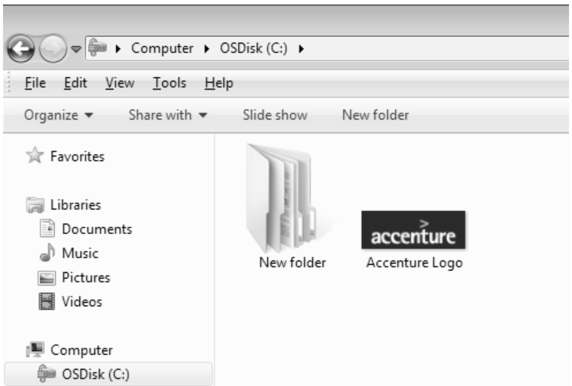


图9-16　添加图像对象时的弹出窗口

可见在添加“图片对象”后，Tableau自动调整了“文本”和“图片”对象的展示宽度，展示高度则与仪表板的默认高度相同（见图9-17）。这是平铺导致的，我们只需要继续在下方添加区域2，自然就会将这部分内容“压小”，并可以手动调整大小。



图9-17　添加“图片对象”后的视图窗口

3. 区域1为公司Logo图像添加网址链接

选中公司Logo“Accenture”，右击鼠标，如图9-18所示，在其中选择“设置URL…”。



图9-18　为图片对象设置网址链接

Tableau自动弹出设置URL窗口，在其中录入需要链接的网址，之后单击“确定”即可（见图9-19）。

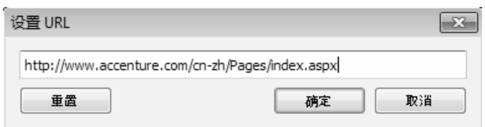


图9-19　输入网址链接

此时单击公司Logo图片即可打开公司官网首页，展示效果见图9-20。

9



图9-20　网址链接添加后鼠标悬停于图片对象上时的展示效果

4. 区域2添加水平布局容器

从“容器与对象窗口”把“水平”字段拖放至“视图窗口”中现有内容的下方灰色阴影处（见图9-21）。

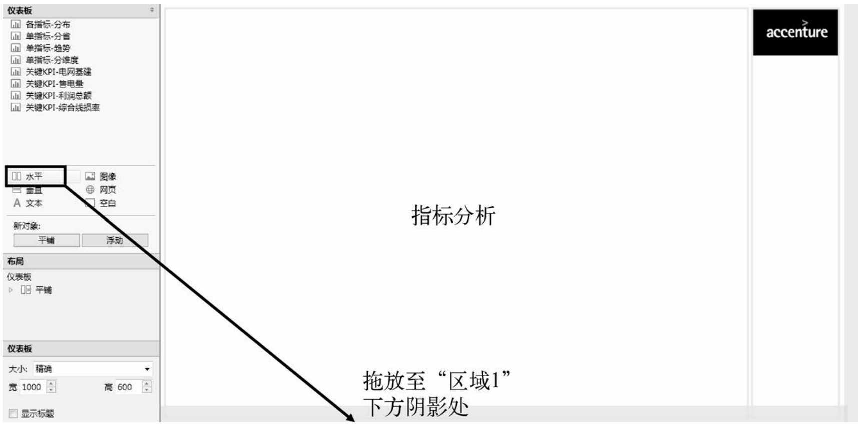


图9-21　添加区域2水平布局容器

此时可选中区域1中文本或对象的下方边缘，调整其高度（见图9-22）。

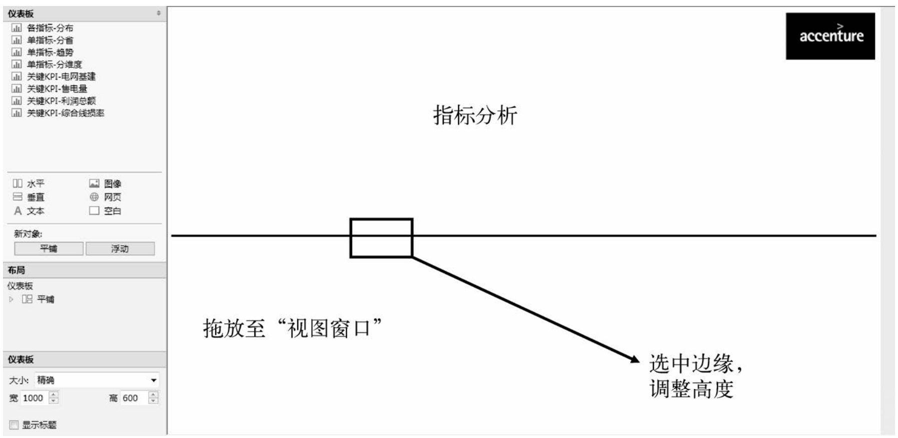


图9-22　可以改变区域1文本或图像的高度

5. 区域2添加关键KPI工作表

从“工作表窗口”把工作表“关键KPI–电网基建”拖放至区域2“水平容器”（见图9-23）。

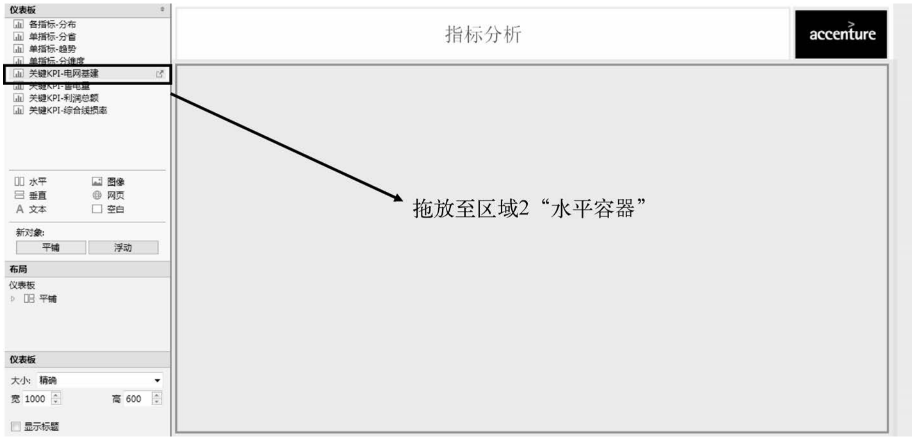


图9-23 区域2添加工作表：“关键KPI–电网基建”

图9-24即是添加“关键KPI–电网基建”后的展示效果，可以发现“电网基建”工作表的筛选器被随机地置于整个仪表板的最右侧。这种展示方式不符合目标仪表板样式，考虑到剩余3个“关键KPI”工作表结构相似，我们全部添加完成后再一并修改。

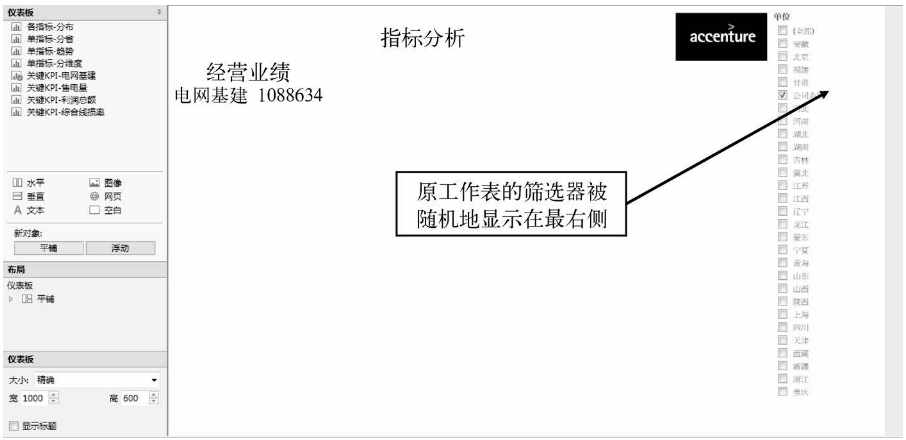


图9-24 添加“关键KPI–电网基建”后的展示效果

从“工作表窗口”把工作表“关键KPI–售电量”、“关键KPI–综合线损率”、“关键KPI–利润总额”拖放至区域2“水平容器”，见图9-25。

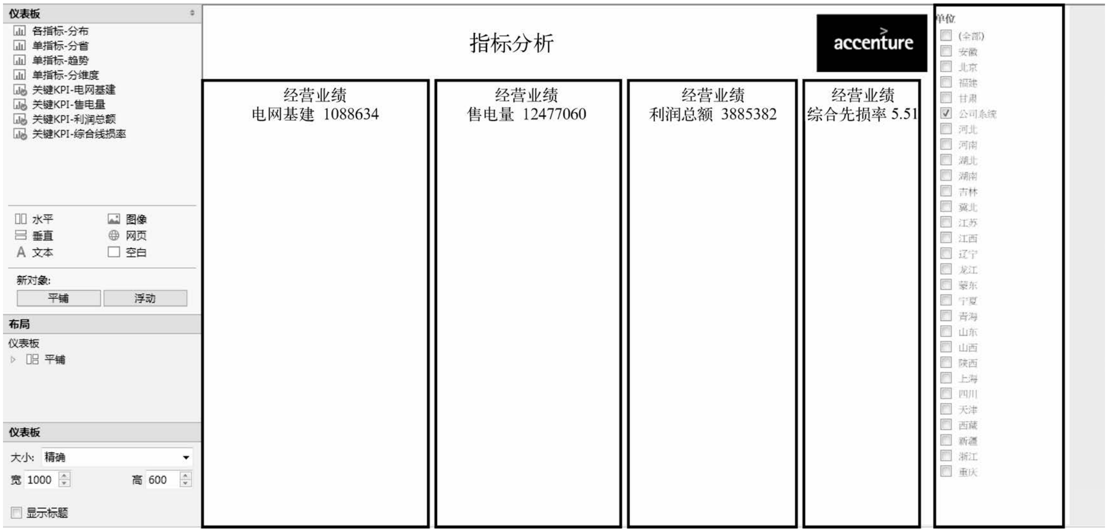


图9-25　添加“关键KPI–售电量”“关键KPI–综合线损率”“关键KPI–利润总额”后的展示效果

6. 区域2显示“统计周期”筛选器

在区域2容器中，选中工作表“经营业绩”，右击或者单击其右上方的下拉箭头，进一步选择“快速筛选器”（见图9-26）➤“统计周期”，展示效果见图9-27。



图9-26　显示“统计周期”筛选器

在此基础上进一步添加文本对象，生成各KPI指标的计量单位，操作方法与步骤4类似，此处不再详细介绍。

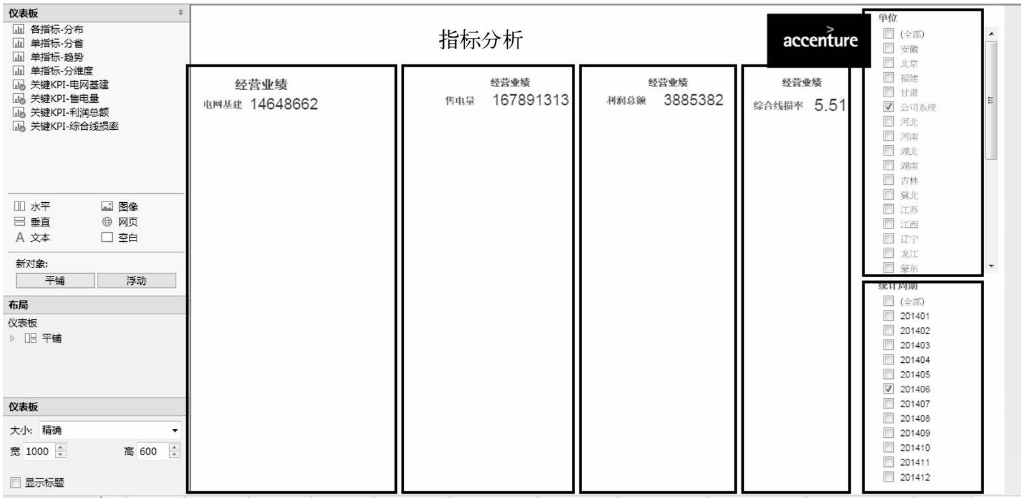


图9-27　显示“统计周期”筛选器后的展示效果

7. 区域2添加垂直布局容器

由于是目前区域2中为水平容器，而目标实例中筛选器应垂直方向上下放置在右侧，所以需要在现有水平容器中内置一个垂直容器（见图9-28）。从“容器与对象窗口”把“垂直”字段拖放至 “视图窗口”区域2中现有内容的最右侧灰色阴影处。



图9-28　添加区域2垂直布局容器

8. 区域2调整筛选器位置

选中筛选器“单位”“统计周期”的上方，将其拖放至区域2的垂直容器中（见图9-29）。

9



图9-29　拖放“单位”“统计周期”筛选器至区域2的垂直容器

9. 区域2调整格式

选中各个“关键KPI”工作表并右击，选择隐藏标题（见图9-30）。

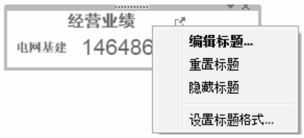


图9-30　隐藏工作表标题

选中各工作表、文本和筛选器的边缘即可左右调整其宽度，此时区域2的布局方式为平铺，展示效果默认为填充至仪表板底部，需在下方添加新的布局容器后才可调整（见图9-31）。

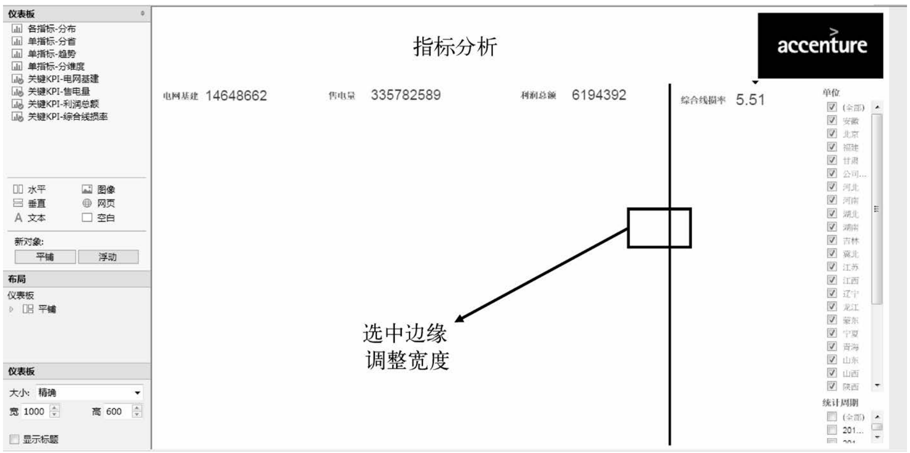


图9-31　调整工作表宽度

选中“单位、统计周期”筛选器，右击切换筛选器展示形式为“单值（下拉列表）”（见图9-32）。



图9-32　切换筛选器展示形式

在经过以上格式调整后，视图窗口展示形式见图9-33。



图9-33　格式调整后视图窗口的展示效果

10. 区域3添加水平布局容器

从“容器与对象窗口”把“水平”字段拖放至区域2下方灰色阴影处（见图9-34）。

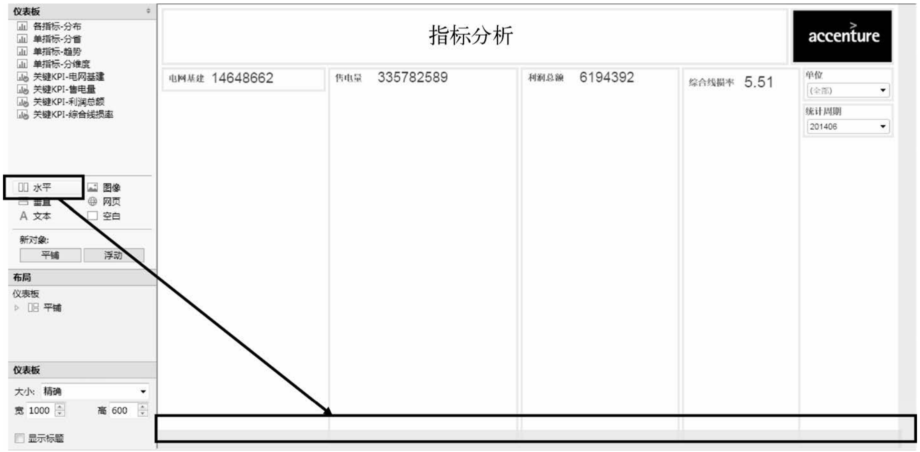


图9-34　添加区域3水平布局容器

11. 区域3添加工作表并调整格式

从“工作表窗口”把工作表“各指标分布情况”“单指标–分省情况”拖放至区域3视图窗口，去除筛选器，显示“形状”图例，设置为“浮动”并调整其位置，然后将工作表调整为合适宽度（见图9-35）。



图9-35　区域3将“形状”图例设置为“浮动”

12. 区域4添加水平布局、工作表并调整格式

从“容器与对象窗口”把“水平”字段拖放至区域3下方灰色阴影处。然后从“工作表窗口”把工作表“各指标月趋势”“单指标–分维度”拖放至区域4视图窗口，去除筛选器，并调整工作表至合适宽度，展示效果见图9-36。

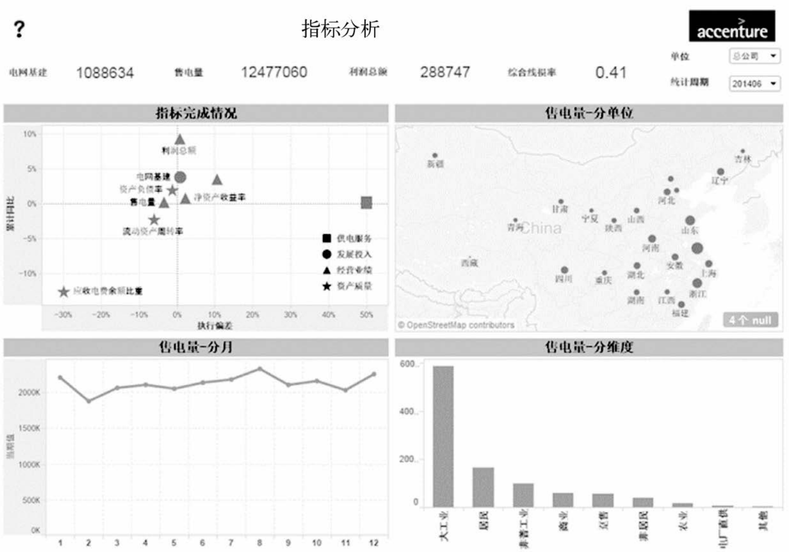


图9-36 区域4添加布局工作表后的展示效果

13. 设置筛选器适用于多个工作表

选中“单位”“统计周期”筛选器，右击或者单击右上角的下拉箭头，选择“应用于工作表”➤“选定工作表”（见图9-37），在弹出窗口中选择“应用于工作表”（即需要适用于工作表）即可。



图9-37 调整筛选器使之适用于多个工作表

14. 通过自定义形状为仪表板添加说明文字

为了在仪表板左上方添加一个问号图标，并显示解释说明文字，首先需要用户自行下载一个

图像文件，如图9-38所示。



图9-38　预先下载的图像文件“问号”

将下载好的文件添加至Tableau的形状文件夹（Shapes）（见图9-39），如果安装过程中未做特殊设置，那么该文件夹的路径为：C:\Users \Documents\My Tableau Repository\Shapes\问号。

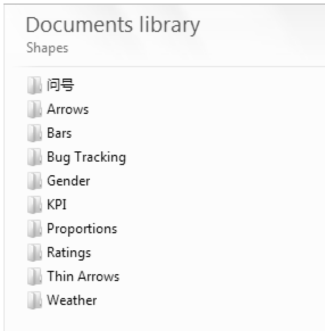


图9-39　添加至Tableau形状文件夹

新建工作表，从度量值窗口任意选取一个度量值拖放至“标记”卡窗口，并将其设置为“形状”，示例见图9-40。



图9-40　添加任一度量值至“标记”卡，并将其设置为“形状”

选中“标记”卡中“形状”选项，单击“更多形状”，在弹出窗口中单击“选择形状板”，可以发现之前添加的图像“问号”，选中并单击“确定”，示例见图9-41。

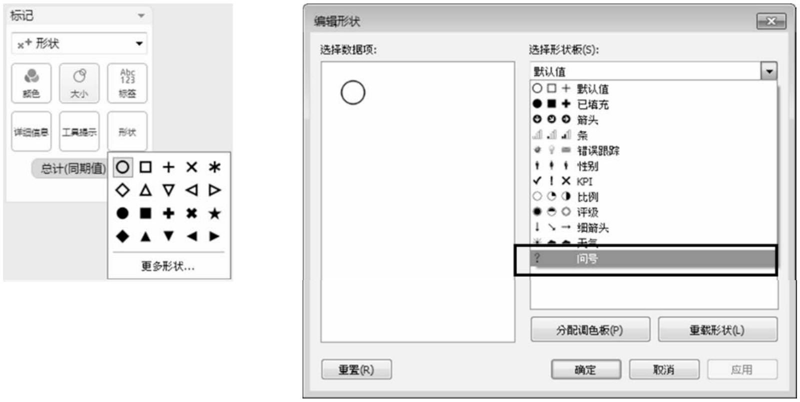


图9-41　选择并应用自定义的图像“问号”

单击“工具提示”选项（见图9-42），在弹出窗口中添加说明性文字（见图9-43），展示效果见图9-44。



图9-42　单击“工具提示”选项

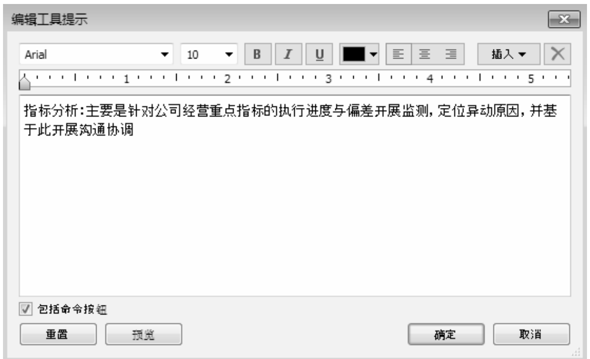


图9-43　在弹出窗口中添加说明性文字

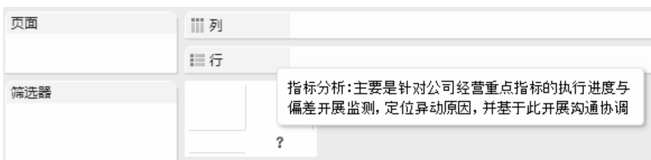


图9-44　添加说明性文字后在工作表中的展示效果

此时在仪表板窗口中即可看到“自定义形状”工作表（见图9-45），用户即可以将其拖放至仪表板区域1来展示总体性介绍，展示效果如图9-46所示。



图9-45　仪表板窗口中的“自定义形状”工作表



图9-46　添加完成后仪表板的展示效果

9.2.3　添加交互操作

可以简单便利地在图表间、视图间添加交互操作是Tableau的关键优势之一，本节基于实例介绍如何向配置好的静态图表间添加“表间筛选”“突出显示”“网址链接”等。

1. 添加表间筛选

● ***确定源工作表、目标工作表及目标筛选器***

筛选器操作可以实现工作表与工作表之间的关联展示以及展示内容的层层钻取。当添加了筛选器操作后，在选中“源工作表”的某个特定对象时，其余的“目标工作表”只展示与选中对象相匹配的内容。如图9-47所示，我们需实现左侧“关键指标完成情况”的钻取查询“单指标–分省情况”。

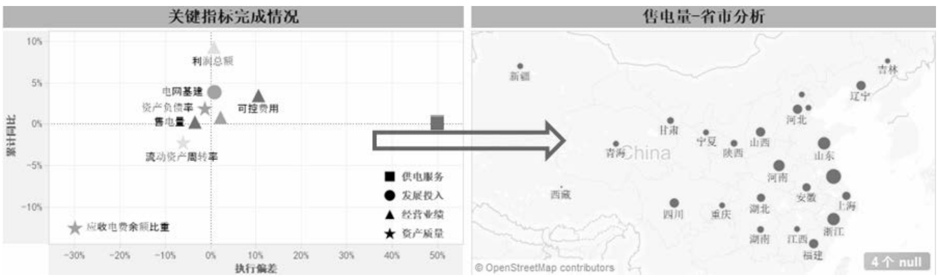


图9-47　确定源工作表、目标工作表

请在仪表板菜单中选择“仪表板”➤“操作”➤“添加操作”➤“筛选器”以添加筛选器操作（见图9-48），并设置相应的源工作表、目标工作表以及目标筛选器。添加筛选器时Tableau的弹出窗口见图9-49。



图9-48 添加交互操作

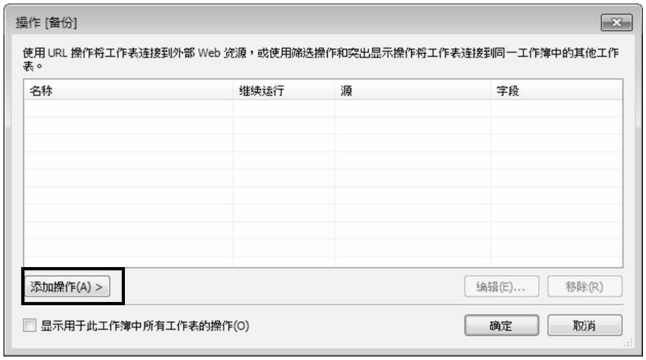


图9-49 添加筛选器时Tableau的弹出窗口

我们可以在“目标筛选器”中设置特定工作表是所有展示字段均受交互操作影响，还是只有特定字段受交互操作影响；若选择“所有字段”，则选择源工作表中的任意字段都可触发交互操作（参见图9-50）。



图9-50 设置交互操作的源工作表与目标工作表

设置目标筛选器的示例见图9-51。



图9-51　设置目标筛选器

此外，我们还可以单击某个工作表右上角的下拉箭头，在弹出菜单中选择“用作筛选器”（见图9-52）以快速添加筛选器操作。“用作筛选器”自动生成的筛选器见图9-53。



图9-52　使用“用作筛选器”

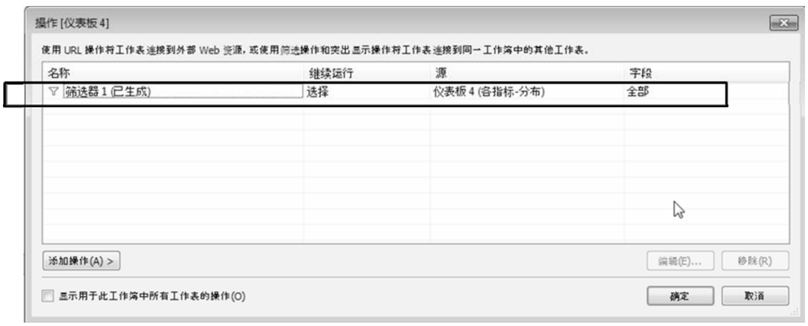


图9-53　“用作筛选器”自动生成的筛选器

“用作筛选器”自动生成的筛选器的默认目标工作表为该仪表板中的全部工作表（见图9-54），需根据实际需求进行相应修改。



图9-54 更新“用作筛选器”自动生成的筛选器

除在同一仪表板的不同工作表间可以配置使用交互操作，Tableau还支持在不同仪表板间进行交互操作。图9-55是“表间筛选”的示例。



图9-55 在选择“目标工作表”时可选择其余仪表板

● ***确定触发筛选方式***

Tableau提供了3种触发筛选的方式，包括“悬停”“选择”“菜单”（见图9-56），其中：① 若选择“悬停”，当鼠标悬停到工作表某个特定对象时交互操作生效；② 若选择“选择”，当鼠标选中工作表某个特定对象时交互操作生效；③ 若选择“菜单”当鼠标悬停或单击工作表某个特定对象时弹出标签卡中会出现对应交互的菜单，单击菜单则该交互操作生效，而且还可以通过右键单击某个特定对象在弹出菜单中选择相应交互操作。详情请参见图9-57 ~ 图9-59。



图9-56　三种触发筛选的方式



图9-57　触发方式为“悬停”，鼠标悬停在左侧某一对象时右侧展示内容即发生变化

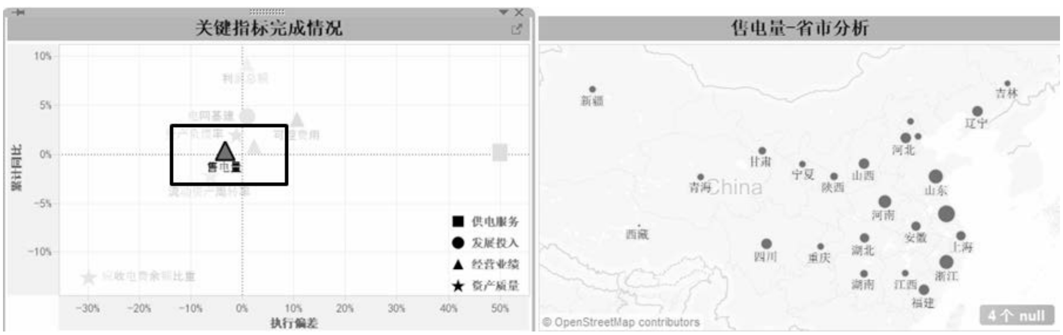


图9-58　触发方式为“选择”，鼠标选择在左侧某一对象时右侧展示内容即发生变化



图9-59 触发方式为“菜单”，鼠标选择左侧某一对象的菜单时右侧展示内容发生变化

- **确定清除筛选后工作表的展现方式**

当取消选择某个工作表上的对象时，对于受其交互操作影响的所有工作表（见图9-60），我们都可以在“添加筛选器操作” ➤ “清除选定内容将会”中设置其展示内容。其中：① 选择“保留筛选器”，则受交互操作影响的所有工作表将在取消选择某个工作表上的对象后仍展示选择该对象时的内容；② 选择“显示所有值”，则受交互操作影响的所有工作表将在取消选择某个工作表上的对象后展示工作表内的所有内容；③ 选择“排除所有值”，则受交互操作影响的所有工作表将在取消选择某个工作表上的对象后不展示任何内容。详情请参见图9-61 ~ 图9-63。



图9-60 设置清除筛选后，受交互操作影响的工作表的展示内容

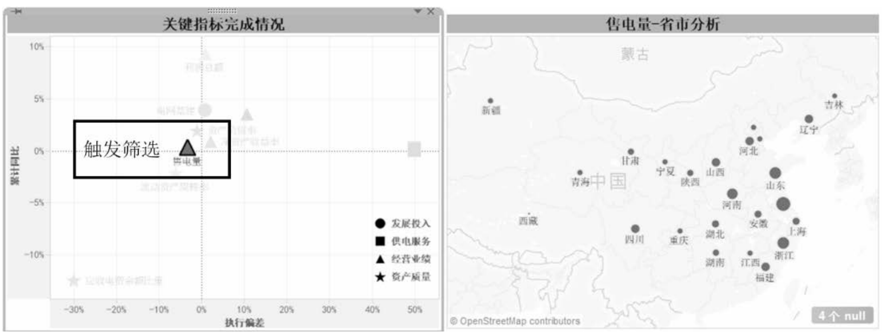


图9-61 选择保留筛选器，则取消选择左侧工作表对象后右侧仍展示交互后的内容

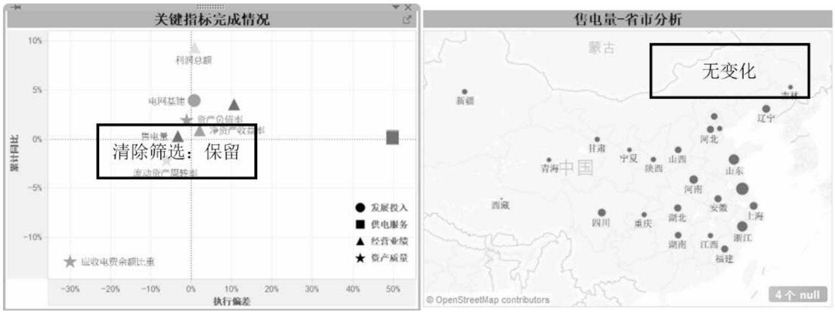


图9-61　（续）

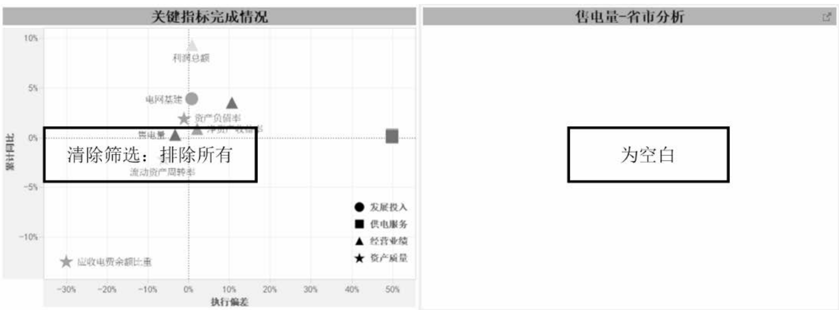


图9-62　选择排除所有值，则取消选择左侧工作表对象后右侧不展示任何内容

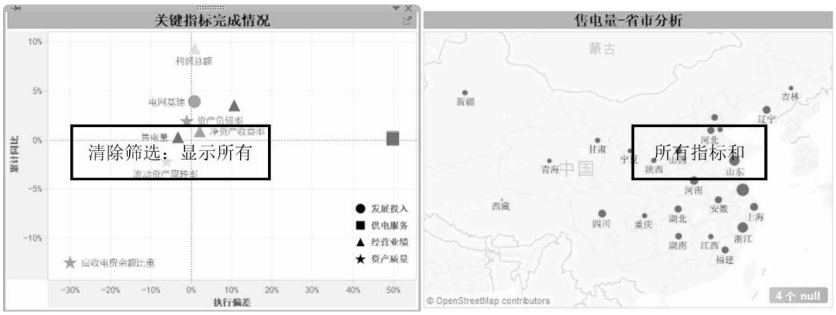


图9-63　选择显示所有值，则取消选择左侧工作表对象后右侧将展示全部内容

注意　在本示例中如果选择“显示所有”，Tableau的实现方式为：展示各省市所有指标、所有月份值的和，可以发现这样展示显然是毫无业务含义的。

2. 添加突出显示

- **确定源工作表、目标工作表（见图9-64）及目标突出显示**

“突出显示”操作可以在“源工作表”的某个特定对象被选中时，高亮显示“目标工作表”

中与选中对象相匹配的内容（见图9-65）。

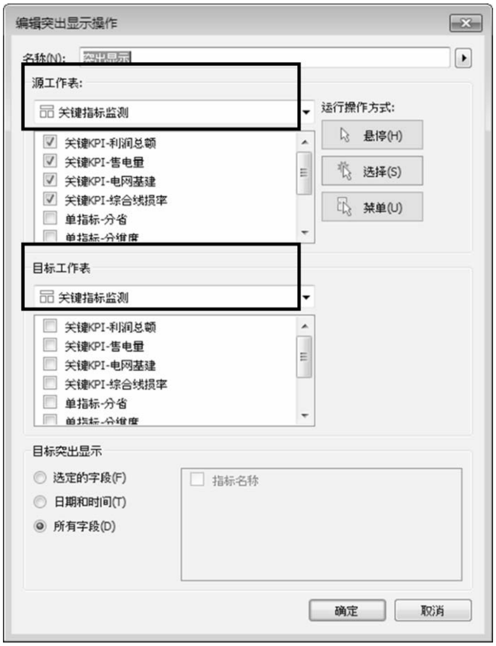


图9-64　设置交互操作的源工作表与目标工作表

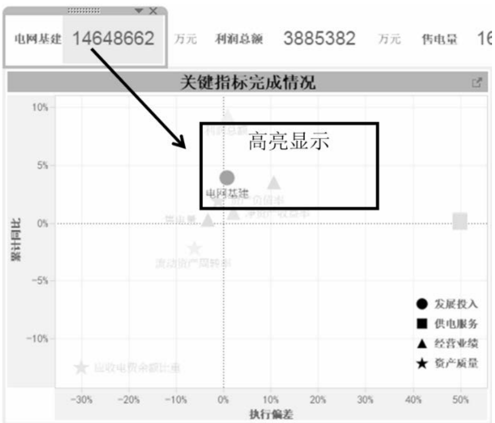


图9-65　选择源工作表中的对象，对应内容在目标工作表中高亮显示

● **确定触发突出显示方式**

与表间筛选相同，Tableau提供了3种触发方式："悬停""选择""菜单"。图9-66给出了选择"菜单"方式时的示例。



图9-66　触发突出显示方式：菜单

3. 添加网址链接

● **确定源工作表、网址链接（见图9-67）。**

添加URL可以使在选中源工作表的特定对象时弹出需要展示的网页。



图9-67　确定源工作表、网址链接

如果想在URL中添加工作表的特定字段作为参数，例如通过搜索引擎展示工作表特定对象的搜索结果，则可以在“编辑URL操作”中单击URL地址栏右侧的三角箭头，并在弹出菜单中选择相应字段。之后，当你单击某个对象触发URL操作时，搜索引擎将自动展示该对象的搜索结果，详见图9-68。

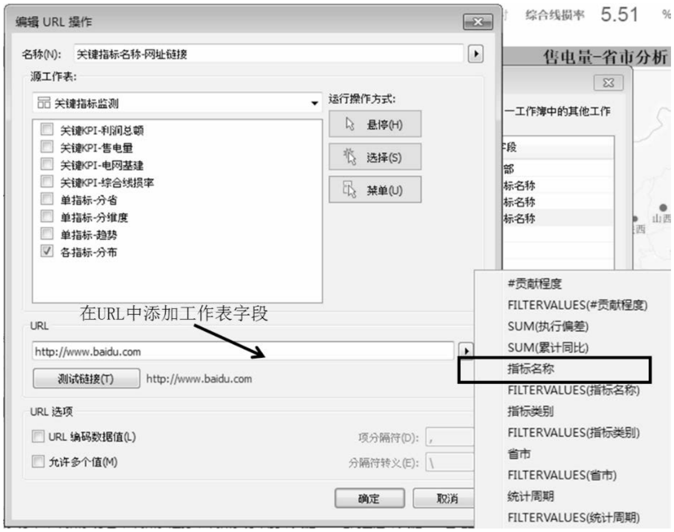


图9-68　在URL中添加工作表字段

注意　需补全URL地址，在本示例中完整的地址为：http://www.baidu.com/s?wd=<指标名称>。

- **确定触发突出显示方式**

Tableau提供了3种触发方式："悬停""选择""菜单"。图9-69给出了"菜单"方式的示例。

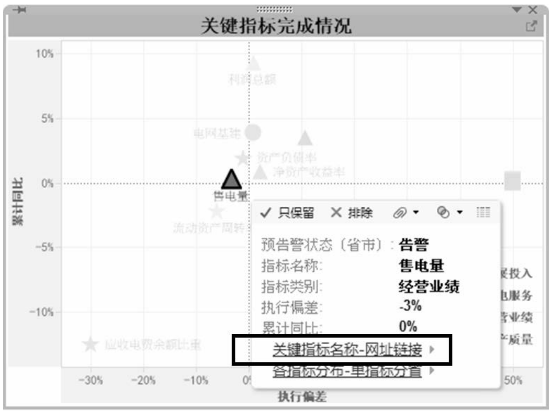


图9-69　选中工作表特定对象可触发URL

注意　当同一对象上有超过两个交互操作时，建议选择设置为"菜单"。如果均为"选择"，则同一次鼠标单击会触发两个交互，展示效果较差。

第 10 章

分析成果共享

10

到本章为止，本书已经讲解了如何连接不同的数据源，以及如何创建简单视图和高级视图进行可视化分析。但目前这些工作成果只能被工作表和仪表板的创建者使用，因此本章将介绍如何让更多的人分享创建者的工作成果。

10.1节介绍如何导出和发布数据及数据源。

10.2节介绍如何把创建的工作表、仪表板和故事保存成图像和打印成PDF。

10.3节介绍多种保存和发布工作簿的方法。所使用的数据源类型不同，保存和发布工作簿的方式也会不同。工作簿发布到Tableau Server服务器和Tableau Online服务器的操作是一致的，所以我们将在10.3.3节一并介绍。

10.1　导出和发布数据（源）

Tableau对于导出一个工作表所使用的部分或者全部数据提供了多种方法，10.1.1节至10.1.3节将展开详细介绍。当然，导出工作簿中所使用的数据源也有多种方式，如导出成.tds文件、.tdsx文件或者.tde文件，详见10.1.4节。有时我们可能需要把不同类型的数据源发布到Tableau 服务器上，以便让更多的人可以查看、使用、编辑或者更新，因此10.1.5节将介绍如何将数据源发布到服务器上。

10.1.1　通过将数据复制到剪贴板导出数据

在视图上右击并在弹出菜单上单击“全选”，或者在视图上右击并在弹出菜单上选择“复制”➤“数据”，或者通过“工作表”➤“复制”➤“数据”，这样将会把视图中的数据复制到剪贴板中。打开 Excel 工作表，然后将数据粘贴到新工作表中即可导出数据，参见图10-1。

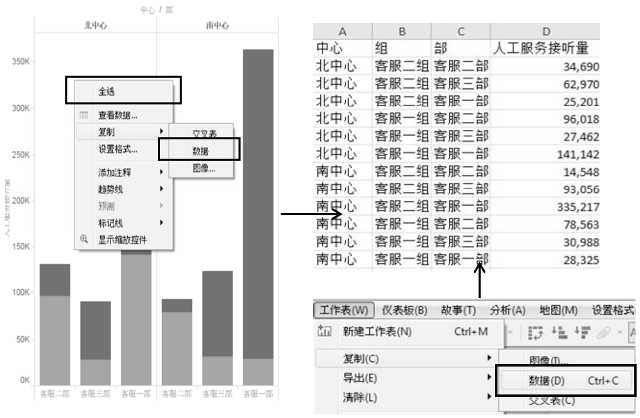


图10-1　通过复制到剪贴板来导出数据

你也可以在视图上右击并在弹出菜单上单击“查看数据”，此时会弹出“查看数据”对话框。在对话框中选择要复制的数据，然后单击对话框右上角的“复制”即会把视图中的数据复制到剪贴板中。打开Excel工作表，然后将数据粘贴到新工作表中，即可导出数据，参见图10-2。

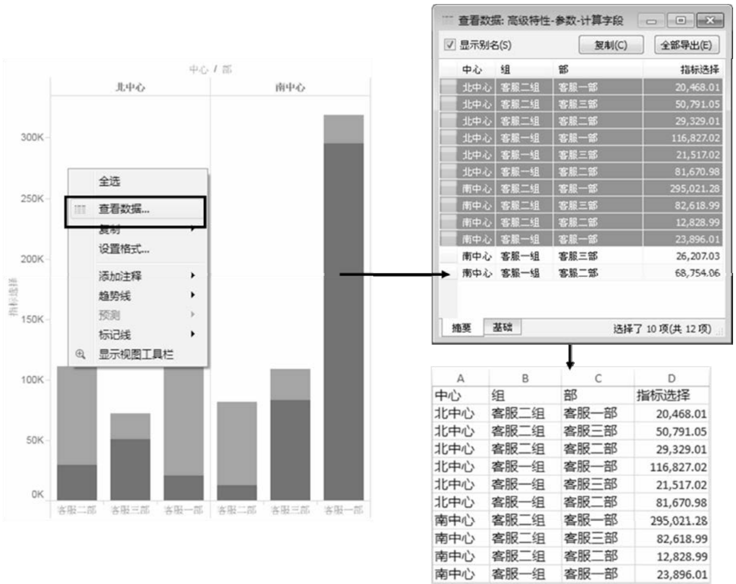


图10-2　通过复制到剪贴板来导出数据

单击“查看数据”对话框右上角的“全部导出”将会打开“导出数据”对话框，请在这里选择一个用于保存导出数据的位置，然后单击“保存”，这样可以把全部数据导出为文本文件（逗号分隔），见图10-3。

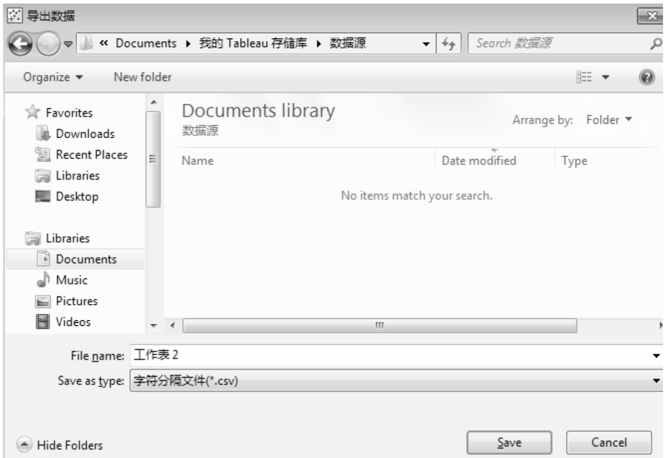
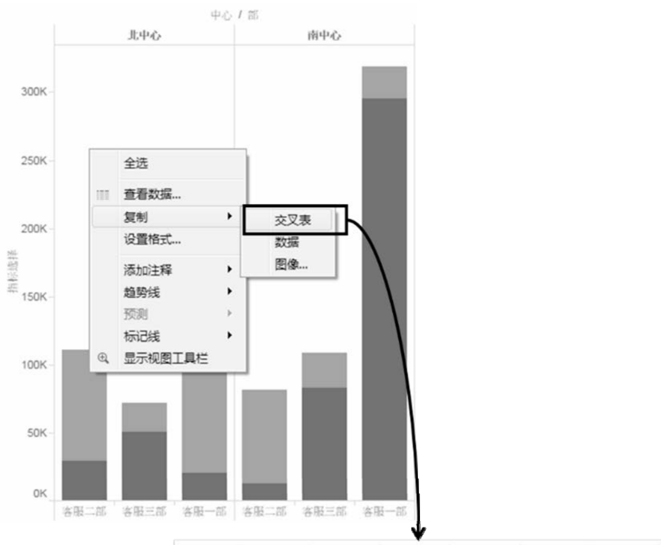


图10-3　导出数据为文本文件（.csv）

你还可以在视图上右击，并在弹出菜单上选择“复制”➤“交叉表”，从而把交叉表（文本表）形式的视图数据复制到剪贴板，见图10-4。然后打开Excel工作表，将数据粘贴到新工作表中，即可导出数据。但我们不能对解聚的数据视图使用此种方法导出数据，因为交叉表是聚合数据视图。换言之，若要使用此方法导出数据，你必须选择“分析”菜单中的“聚合度量”选项。



A	B	C	D	E	F	G
	中心 / 部	中心 / 部	中心 / 部	中心 / 部	中心 / 部	中心 / 部
	北中心	北中心	北中心	南中心	南中心	南中心
组	客服二部	客服三部	客服一部	客服二部	客服三部	客服一部
客服一组	81,671	21,517	116,827	68,754	26,207	23,896
客服二组	29,329	50,791	20,468	12,829	82,619	295,021

图10-4　复制交叉表来导出数据

10.1.2　以Access数据库文件导出数据

我们也可以以Access数据库文件的方式导出当前工作表中的数据，方法是选择“工作表”➤“导出”➤“数据”，在弹出对话框中为待导出的Access数据库文件指定存放路径和文件名（Access数据库的文件扩展名为.mdb），见图10-5。

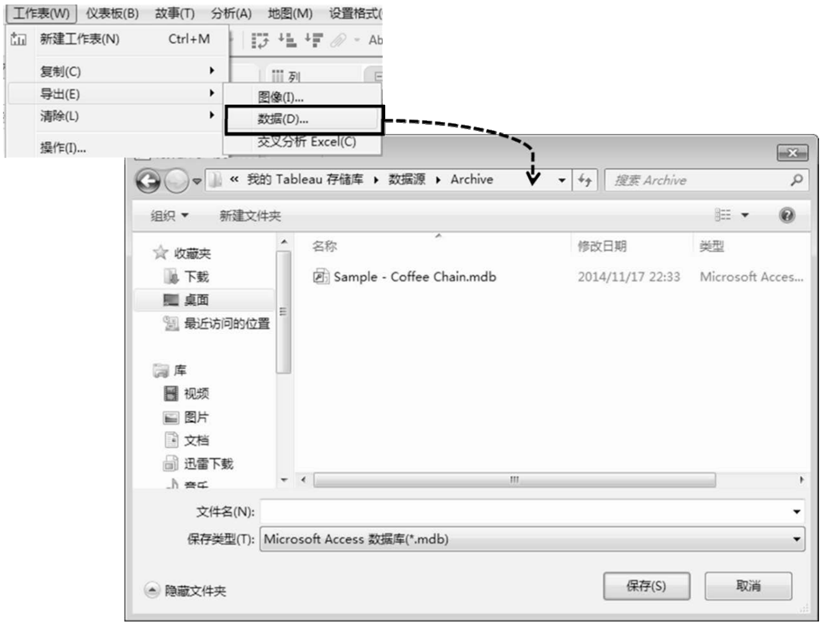


图10-5　为Access数据库文件指定存放路径和名称

在图10-5中指定路径并键入数据库名称，单击“保存”将显示“将数据导出到Access”对话框。如果选择“导出后连接”选项，你可以立即连接到新数据源并继续使用Access，而不会中断工作。单击“确定”即把当前视图中的数据导出为Access数据库文件，见图10-6。



图10-6　把数据导出为Access数据库文件

10.1.3　以交叉分析（Excel）方式导出数据

选择菜单栏中“工作表”➤“导出”➤ “交叉分析Excel”，Tableau将自动创建一个Excel文件，并把当前视图中的交叉表数据粘贴到这个新的Excel工作簿中，见图10-7。

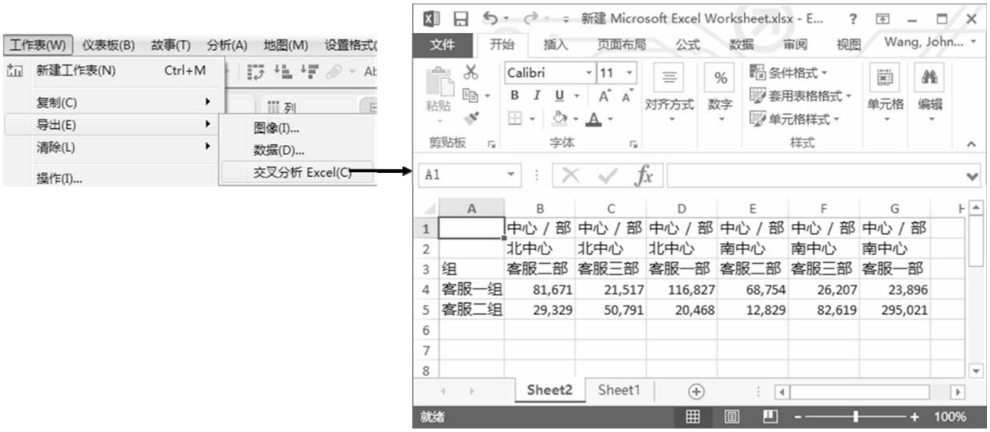


图10-7　把数据导出为Excel文件

说明　将交叉表复制到Excel更为直接，但由于它会带格式复制数据，因此可能会降低性能。如果需要导出的视图包含大量数据，你会看到一个对话框，要选择是否复制格式设置选项，如果选择不复制格式则可以提高性能。此外，你不能对解聚的数据视图使用此种方法导出数据，因为交叉表是聚合数据视图。

10.1.4　导出数据源

有两种方法可以将所有数据或数据子集导出到新数据源：第一种方法是在“数据”菜单上选择数据源，然后选择“添加到已保存的数据源”来导出数据源；第二种方法是使用 Tableau 数据提取导出数据源。第一种方法将会以数据源（.tds）文件或打包数据源（.tdsx）文件保存数据，而第二种方法创建的是数据源的已保存子集（.tde）文件，可用于提高性能，还可提供对数据的脱机访问，从而进行脱机分析。

1. 利用“添加到已保存的数据源”导出数据源

通过“数据”➤“<数据源名称>”➤“添加到已保存的数据源”可以导出数据源文件（.tds）和打包数据源文件（.tdsx），使用这种方式导出的数据源不必在每次需要使用该数据源时都创建新连接。因此，如果经常多次连接同一数据源，我们推荐用这种方式导出数据源（参见图10-8）。

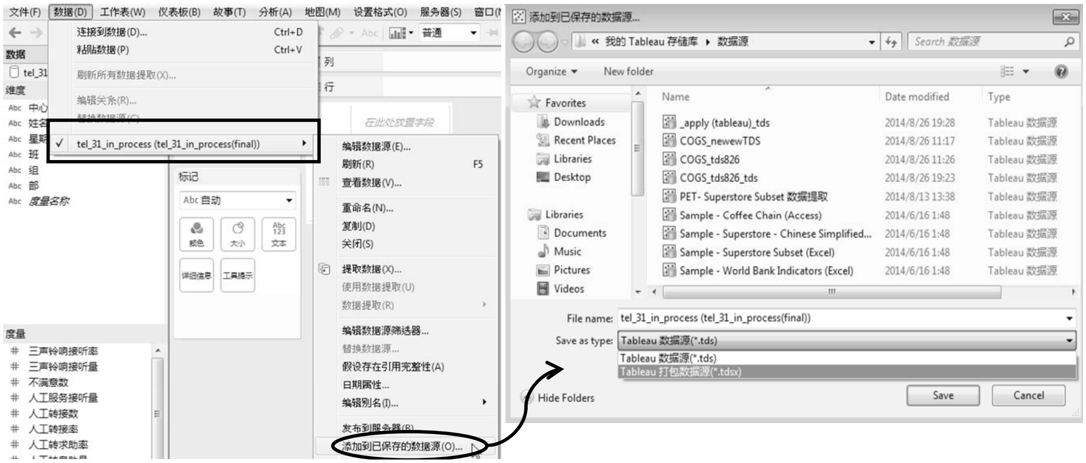


图10-8　利用“添加到已保存的数据源”导出数据源

在弹出的“添加到已保存的数据源”对话框上，请选择一个用于保存数据源文件的位置。默认情况下，数据源文件存储在Tableau存储库的数据源文件夹中。如果不更改存储位置，新.tds或 .tdsx 文件将在开始页面中的“数据”区域中的“已保存数据源”部分中列出（参见图10-9）。



图10-9　把数据源导出到Tableau存储库后数据源显示在开始页面中

由图10-8可以看出，我们可采用以下两种格式来导出数据源。

- **数据源（.tds）**。如果连接的是本地文件数据源（Excel、Access、文本、数据提取），导出的数据源文件（.tds）包含数据源类型和文件路径。如果连接的是实时数据源，导出的数据源文件（.tds）包含数据源类型和数据源连接信息（服务器地址、端口、账号）。无论连接到本地文件还是数据库服务器数据源，数据源文件（.tds）都还包括数据源的默认属性（数字格式、聚合方式和排序顺序等）和自定义字段（如组、集、计算字段和分级字段）。

10

- **打包数据源（.tdsx）**。如果连接的是本地文件数据源（Excel、Access、文本、数据提取），导出的打包数据源文件（.tdsx）不但包含数据源文件（.tds）中的所有信息，还包含本地文件数据源的副本，因此可与无法访问你计算机上本地存储的原始数据的人共享.tdsx数据源。如果连接的是实时数据源，采用打包数据源（.tdsx）和数据源（.tds）两种格式所导出文件包含的内容完全相同。

说明　如果创建了参数，并在自定义字段时使用了参数，之后使用“添加到已保存的数据源”方式导出数据源文件（.tds或.tdsx），数据源文件中将包含创建的参数；如果仅仅创建了参数，但没有被自定义字段使用，之后使用“添加到已保存的数据源”方式导出数据源文件（.tds或.tdsx），数据源文件中将不包含创建的参数。

打包数据源.tdsx文件类型是一个压缩文件，可用于与无法访问你计算机上本地存储的原始数据的人共享数据源。

---

2. 利用“数据提取”导出数据源

通过“数据”➤“<数据源名称>”➤“提取数据”打开“提取数据”对话框。在对话框中，我们可以定义筛选器来限制将提取的数据，也可以指定是否聚合数据来进行数据提取（如果对数据进行聚合可以最大限度地减小数据提取文件的大小并提高性能，如按照月度聚合数据），还可以选定想要提取的数据行数，或者指定数据刷新方式（增量刷新或者完全刷新），完成后请单击“数据提取”。在随后显示的对话框中你要选择一个用于保存提取数据的位置，然后为该数据提取文件指定文件名称，最后单击“保存”便可创建数据提取文件（.tde）并完成数据源的导出。3.4.1节详细介绍了数据提取方法，这里不再赘述。

用这种方式导出数据源有很多好处：可以避免频繁连接数据库，从而减轻数据库负载；若进行包含数据样本的数据提取，在制作视图时，不必在每次将字段放到功能区上时都执行耗时的查询，因而可以提高性能；在不方便新建数据源服务器时，数据提取可提供对数据的脱机访问，进行脱机分析；而且当基础数据发生改变时，还可以刷新提取数据，与数据库服务器端的数据保持一致。

说明　使用数据提取方式导出的数据源文件（.tde），包括数据源类型、数据源连接信息、默认属性（数字格式、聚合方式和排序顺序等）和自定义字段（如组、集、计算字段和分级字段），但不包含参数。如果创建自定义字段时使用了参数，并且之后进行了数据提取，那么再使用提取数据时，使用了参数的自定义字段将变成无效字段。

10.1.5　发布数据源

上一节我们介绍了如何把数据源导出到本地，你还可以将本地文件数据源或实时连接的数据

库数据源发布到Tableau Online服务器或Tableau Server服务器。本节仅介绍如何将数据源发布到Tableau Server，发布到Tableau Online服务器上的方法与其类似。

在“数据”菜单上选择数据源，然后选择“发布到服务器”。如果尚未登录 Tableau Server，则会弹出“Tableau Server 登录”对话框，请在对话框中输入服务器名称或URL、用户名和密码（参见图10-10）。



图10-10　连接Tableau Server服务器

成功登录Tableau Server服务器后会看到“将数据源发布到 Tableau Server”对话框（参见图10-11）。在对话框中需要指定以下几项内容。

- **项目**。一个项目就像是一个可包含工作簿和数据源的文件夹，在Tableau Server上创建。Tableau Server 自带一个名为 “默认值”的项目，所有数据源都必须发布到项目中。
- **名称**。在“名称”文本框中提供数据源的名称。使用下拉列表选择服务器上的现有数据源，使用现有数据源名称进行发布时，服务器上的数据源将被覆盖。发布者必须具有“写入/另存到Web”权限才能覆盖服务器上的数据源。
- **身份验证**。如果数据源需要用户名和密码，则可以指定在将数据源发布到服务器上时应如何处理身份验证。可用选项取决于所发布的数据源的类型：当发布的数据源是本地文件时，身份验证只有“无”选项；当发布数据提取数据源时，身份验证有“无”和“嵌入式密码”两个选项；当发布的数据源是实时新建数据源时，身份验证有“提示用户”和“嵌入式密码”两个选项。
- **添加标记**。可以在“标记”文本框中键入一个或多个描述数据源的关键字。在服务器上浏览数据源时，标记可帮助查找数据源。各标记应通过逗号或空格来分隔，如果标记中包含空格，则键入该标记时应将其放在引号中（如“Profit Data”）。

10

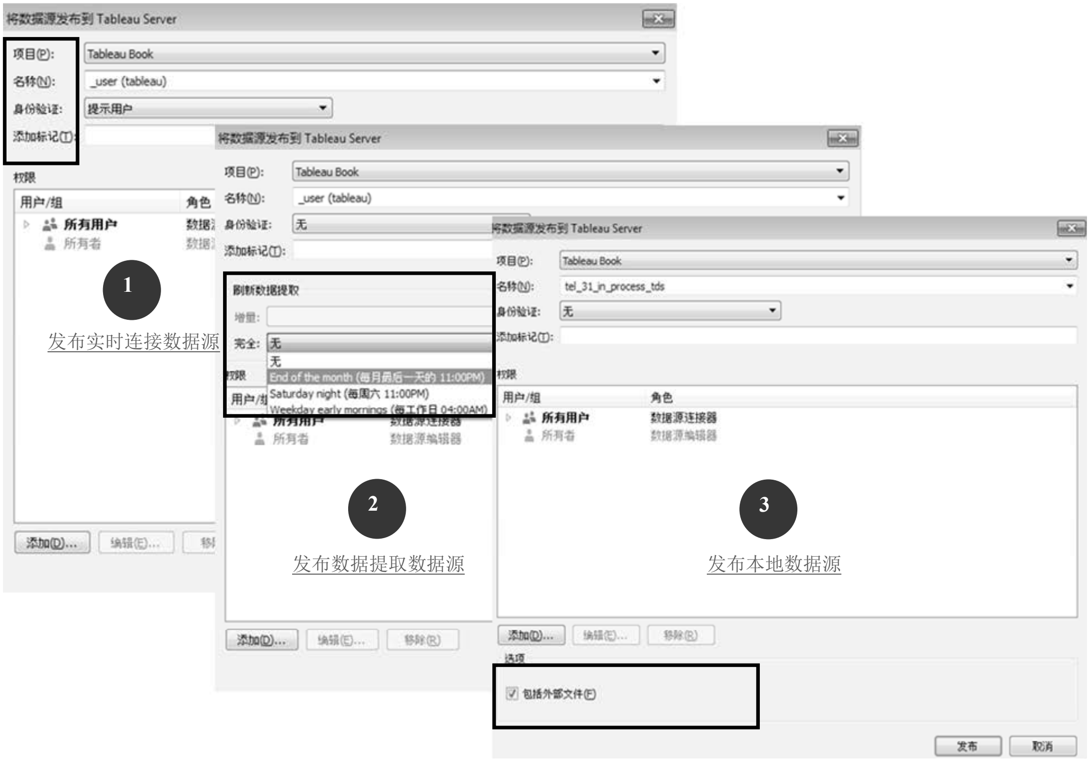


图10-11　将数据源发布到Tableau Server

所发布的数据源的类型不同，“将数据源发布到Tableau Server”对话框中的选项也会略有差异。

(1) 发布实时提取数据源：除了以上4项内容，不需要指定额外的内容。

(2) 发布数据提取数据源：除了以上4项内容，还需使用下拉列表指定刷新数据提取的频率。

(3) 发布本地文件数据源：除了以上4项内容，还需选择是否“包含外部文件”。如果选择“包含外部文件”将会发布数据源的副本。如果不选择“包含外部文件”，则其他人无法联机查看数据源。

10.2　导出图像和 PDF 文件

本节将介绍如何导出Tableau页面，通过复制图像、导出图像以及打印为PDF这3种方式我们可将Tableau动态交互文件转换为打印的静态文件。

10.2.1　复制图像

在工作表工作区环境下，选择“工作表”➤“复制”➤“图像”，并在弹出的“复制图像”

对话框中选择要包括在图像中的内容以及图例布局（如果该视图包含图例），然后单击“复制”，此时Tableau会将当前视图复制到剪切板中（参见图10-12）。

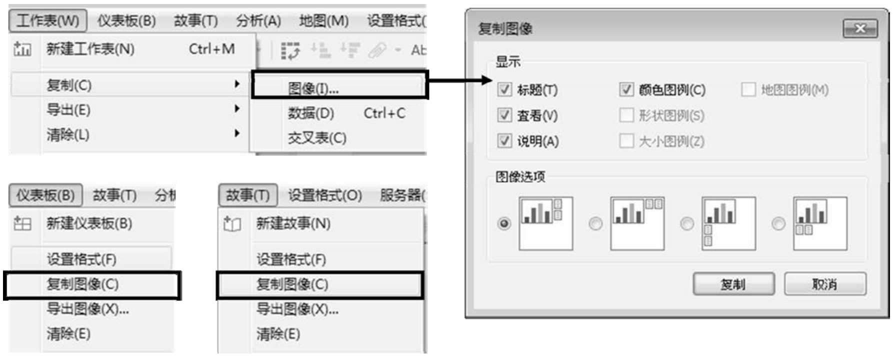


图10-12　复制图像

在仪表板工作区环境下选择“仪表板”➤“复制图像”，或者在故事工作区环境下选择“故事”➤“复制图像”，可以将仪表板中的整个视图或故事中当前故事点的整个视图复制到剪贴板中。用这两种方法复制图像均不会弹出“复制图像”对话框。

把视图复制至剪贴板中后，你可以打开目标应用程序，然后从剪贴板粘贴。

10.2.2　导出图像

选择菜单栏中的“工作表”➤“导出”➤“图像”，并在弹出的“导出图像”对话框中选择要包括在图像中的内容以及图例布局（如果该视图包含图例），然后单击“保存”，此时弹出“保存图像”对话框（参见图10-13）。



图10-13　导出图像

你还可以在仪表板工作区环境下选择“仪表板”➤“导出图像”，或者在故事工作区环境下选择“故事”➤“导出图像”，同样会看到“保存图像”对话框（参见图10-14）。

导出图像与复制图像不同，导出图像会弹出“保存图像”对话框，在对话框中你可以对导出图片的类型（如jpg、png、bmp等）、名称和路径进行设置。



图10-14　保存图像

10.2.3　打印为PDF

选择菜单栏中的“文件”➤“打印为PDF”，并在弹出的“打印为PDF”对话框中单击“确定”，这样可以将一个视图、一个仪表板、一个故事或者整个工作簿发布为PDF（参见图10-15）。



图10-15　打印为PDF

我们通过“打印为PDF”对话框选择和设置以下选项。

- ❑ 打印范围设置：选择“整个工作簿”选项将把工作簿中的所有工作表发布为PDF，选择“当前工作表”将仅发布工作簿中当前显示的工作表，选择“选定工作表”选项仅发布选定的工作表。
- ❑ 纸张尺寸选择：可以利用“纸张尺寸”下拉菜单选择打印纸张大小。如果“纸张尺寸”选择为“未指定”，则纸张尺寸将扩展至能够在一页上放置整个视图的所需大小。
- ❑ 其他选项：如果选中“打印后查看PDF文件”选项，创建PDF后将自动打开文件，但请注意只有在计算机上安装了Adobe Acrobat Reader或Adobe Acrobat时才会提供此选项。如果选中“显示选定内容”选项，视图中的选定内容将保留在PDF中。

说明 (1) 打印工作表时，不包含快速筛选器。若要显示快速筛选器，请创建一个包含工作表的仪表板，并将该仪表板打印为PDF。
(2) 在将仪表板打印为PDF时，不会包含网页对象的内容。
(3) 在将故事打印为PDF时，将把故事中的所有故事点都发布为PDF。

10.3　保存和发布工作簿

本节介绍如何保存配置好的Tableau文件，以及如何将Tableau内容发布到服务器进行成果共享和发布。

10.3.1　保存工作簿

工作簿是工作表的容器，用于保存创建的工作内容，由一个或多个工作表组成。在打开Tableau Desktop应用程序时，Tableau会自动创建一个新工作簿。选择“文件”➤“保存”或按Ctrl＋S会弹出“另存为”对话框（首次保存才会弹出），其中要指定工作簿的文件名和保存路径（参见图10-16）。

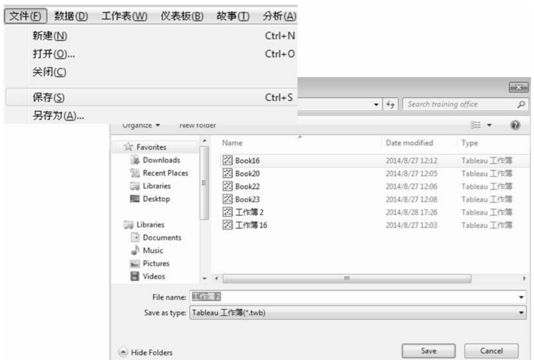


图10-16　保存工作簿

默认情况下，Tableau使用.twb扩展名来保存文件，默认位置为 Tableau 存储库中的工作簿文件夹。不过，你可以选择将Tableau工作簿保存到任何其他目录。

Tableau 文件名不得包含以下字符：正斜杠（/）、反斜杠（\）、大于号（>）、小于号（<）、星号（*）、问号（?）、双引号（"）、竖线符号（|）、冒号（:）或分号（;）。若要另外保存已打开工作簿的副本，请选择“文件”➤“另存为”，然后用新名称保存文件。

10.3.2　保存打包工作簿

保存成工作簿文件时也将保存指向数据源和其他一些资源（如背景图片文件、自定义地理编

码文件）的链接，下次打开该工作簿时将自动使用相关数据和资源来生成视图。这是大多数情况下的工作簿保存方式。但是，如果想要与无法访问所使用数据和资源的其他人共享工作簿，你可以把制作好的工作簿以打包工作簿的形式保存。

Tableau使用.twbx扩展名来保存打包工作簿文件，文件中包含本地文件数据源（Excel、Access、文本、数据提取等文件）的副本、背景图片文件和自定义地理编码。保存打包工作簿的方式有如下两种。

方式1：在菜单中选择“文件”➤“另存为”，在弹出的“另存为”对话框中指定打包工作簿的文件名，并在“保存类型”下拉列表中选择“Tableau打包工作簿（.twbx）”，最后单击“保存”（参见图10-17）。



图10-17　保存打包工作簿（方法1）

方式2：在菜单中选择“文件”➤“导出打包工作簿”，在弹出的“导出打包工作簿”对话框中指定打包工作簿的文件名，最后单击“保存”（参见图10-18）。



图10-18　保存打包工作簿（方法2）

> **说明** 打包工作簿文件（.twbx）类型是一个压缩文件，你可以在Windows资源管理器中的打包工作簿文件上右击，然后选择“解包”。将工作簿解包后会看到一个普通工作簿文件（.twb）和一个文件夹，该文件夹包含与该工作簿一起打包的所有数据源和资源。

10.3.3 将工作簿发布到服务器

通过发布工作簿可将工作成果发布到Tableau服务器上，如Tableau Server服务器和Tableau Online服务器。工作簿发布到Tableau Server和Tableau Online的操作是一致的，区别在于发布的目的地不同，及对数据源的类型要求略不同，接下来我们将详细介绍。

发布工作簿时可以将其添加到服务器上的指定项目下，隐藏某些工作表，添加标记以增强可搜索性，指定权限以控制对服务器上工作簿的访问，以及选择嵌入数据库密码以便在 Web 上进行自动身份验证。

在“服务器”菜单上选择数据源，然后选择“发布工作簿”。如果尚未登录 Tableau服务器，你会看到“Tableau Server 登录”对话框。请在对话框中输入服务器名称或URL、用户名和密码，然后单击“登录”，见图10-19。

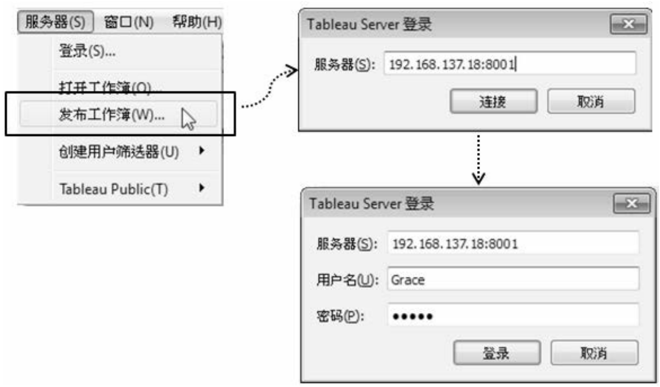


图10-19 登录Tableau服务器

成功登录Tableau服务器后，你会看到“将工作簿发布到 Tableau Server”对话框。所发布的工作簿中使用的数据源的类型不同，对话框中的选项也会略有差异。但无论发布的工作簿使用何种数据源，均需要指定以下5项内容。

- **项目**。一个项目就像是一个可包含工作簿和数据源的文件夹，在Tableau Server上创建。Tableau Server 自带有一个名为“默认值”项目，所有工作簿都必须发布到项目中。
- **名称**。在“名称”文本框中提供工作簿的名称。使用下拉列表选择所选项目下的现有工作簿，使用现有工作簿名称进行发布时，服务器上的工作簿将被覆盖。发布者必须具有“写入/另存到 Web”权限才能覆盖服务器上的工作簿。

- **添加标记**。可以在“标记”文本框中键入一个或多个描述工作簿的关键字。在服务器上浏览工作簿时，标记可帮助查找工作簿。各标记应通过逗号或空格来分隔，如果标记中包含空格，则键入该标记时应将其放在引号中（如“Profit Workbook”）。
- **查看权限**。发布者可以指定相应权限来允许或拒绝对服务器上工作簿的访问。有关权限设置和分配的详细信息，请参见11.3节中介绍的服务器安全机制。
- **要共享的视图**。在此选项窗口中可以选择要在Tableau Server 上共享的工作表。任何未选择的工作表都在服务器上隐藏，当要发布仪表板或故事而不是发布用于创建该仪表板或故事的工作表时，显示和隐藏工作表十分有用。但是拥有“下载/另存到Web”权限的任何人都可从服务器下载工作簿，然后访问隐藏的工作表。因此，隐藏工作表不是隐藏信息的安全方法。

如果所发布的工作簿中使用的数据源的类型不同，“将工作簿发布到 Tableau Server”对话框中除以上5个选项外的其他选项会略有差异。

1. 使用本地文件数据源

如果发布的工作簿连接的是本地文件数据源（Excel、Access和文本文件等），弹出的“将工作簿发布到 Tableau Server”对话框见图10-20。



图10-20　将连接本地文件数据源的工作簿发布到Tableau服务器

如果在图10-20中选择“包含外部文件”，发布的将是.twbx打包工作簿文件，它会包含本地数据源、背景图像和自定义地理编码的副本；如果不选择“包含外部文件”，发布的将是.twb工作簿文件，其他人无法联机查看工作簿；如果不选择“包含外部文件”而单击“发布”，你将看到如图10-21所示的错误提示框。

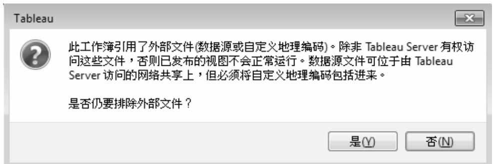


图10-21　关于排除外部文件的错误提示框

2. 使用实时新建数据源

如果发布的工作簿连接的是实时数据源，发布到服务器上的将是.twb工作簿文件，弹出的“将工作簿发布到Tableau Server”对话框见图10-22。



图10-22　将连接实时数据源的工作簿发布到Tablea服务器上

可以看出，图10-22相比图10-20多出了一个“身份验证”按钮；单击“身份验证”将弹出“身份验证对话框”。在对话框中，你可以对数据源的身份验证进行设置，选项包括“提示用户”和“嵌入式密码”：如果选择“嵌入式密码”，用户登录服务器后无需输入所连接数据源的数据库上的账号即可查看此工作簿；如果选择“提示用户”，用户登录服务器后需要输入所连接数据源的数据库上的账号才可查看此工作簿。

说明　发布到Tableau Online时所有的数据源都要进行数据提取，不允许实时连接数据源，而发布到Tableau Server上的工作簿则没有这一限制。

3. 使用数据提取数据源

如果发布的工作簿连接的是数据提取数据源，发布到Tableau服务器上的将是.twbx打包工作簿文件。“将工作簿发布到 Tableau Server”对话框见图10-23。



图10-23　将使用数据提取数据源的工作簿发布到Tableau 服务器上

可以看出，图10-23相比图10-22和图10-20多出了一个“计划与身份验证”按钮。单击该按钮将看到“计划与身份验证”对话框，在对话框中可以对数据源的数据提取计划和身份验证进行设置。服务器提供了几种不同的数据提取更新计划，你可以根据需要任选一种。对于身份验证，与发布连接实时数据源的工作簿不同，这里的可选项是“无”和“嵌入式密码”：如果选择数据源的身份验证为“无”，则用户无法在服务器上刷新此工作簿；如果选择数据源的身份验证为“嵌入式密码”，在用户登录服务器后，无需输入所连接数据源的数据库上的账号即可查看此工作簿。

说明　如果发布到服务器上的工作簿连接的数据源是基于本地文件提取的数据，则无法对此数据源进行“身份验证”设置。

10.3.4　将工作簿保存到Tableau Public上

除了可以把工作簿发布到Tableau Server和Tableau Online服务器，我们还可以把工作簿保存到由Tableau托管的免费且公开的服务器Tableau Public上。保存到Tableau Public的工作簿的数据不得

超过100万行，且我们无法把连接到实时数据源的工作簿保存到Tableau Public。如果尝试把连接到实时数据源的工作簿保存到Tableau Public上，Tableau 会自动提取数据。

选择菜单栏“服务器”➤“Tableau Public”➤“保存至Web”，见图10-24。



图10-24　把工作簿保存到Tableau Public上

如未登录到服务器，你会看到Tableau Public登录对话框，输入Tableau Public账号名和密码即可登录。如果未注册过Tableau Public账号，在登录对话框中选择“Create one for FREE!”可以免费创建一个，见图10-25。

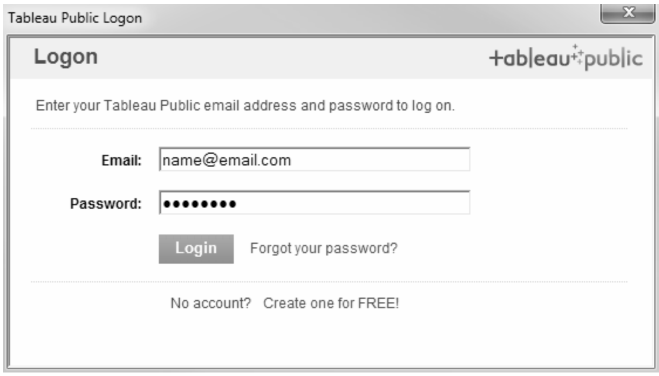


图10-25　登录Tableau Public服务器

登录成功之后此工作簿会被保存到Tableau Public服务器上，保存成功后将显示已发布的工作簿，见图10-26。该页面允许预览所有已保存的工作表。选择一个工作表并单击视图左下角的“共享”按钮可以获得一个链接，你可将此链接通过电子邮件发送给他人，或者把它嵌入到网页中。

10

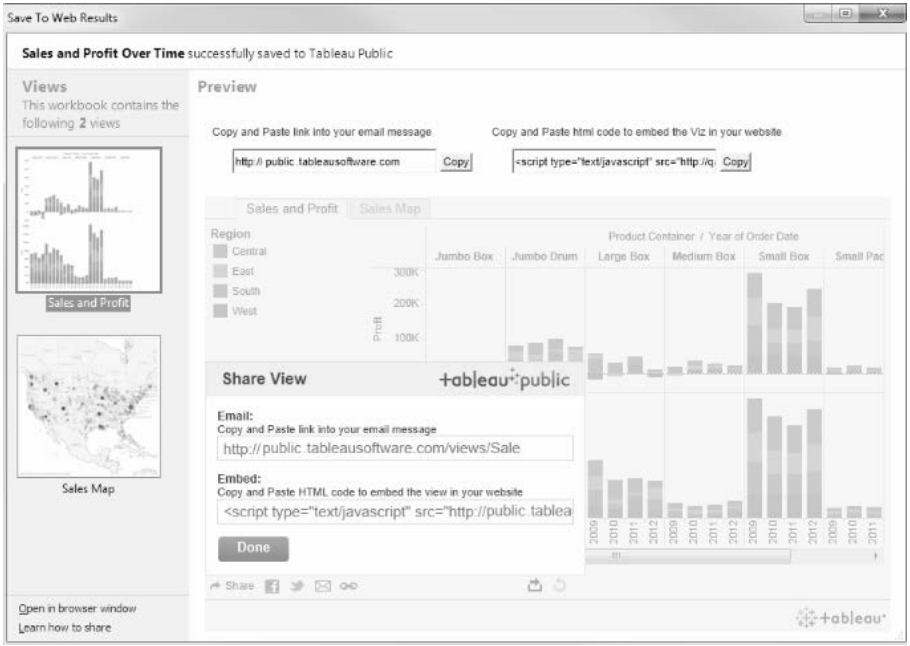


图10-26　成功保存到Web

说明　执行本节介绍的方法也可将工作簿发布到Tableau Public Premium，请确保在登录时使用Tableau Public Premium用户名和密码。

保存到Tableau Public的工作簿和基础数据是公开可用的，你可在网址 http://public.tableausoftware.com 来联机管理保存到Web上的内容。

本章主要介绍了如何把创建的工作簿以及使用的数据源保存在本地或与他人通过网络共享。

你可以将使用的数据源通过很多种方式导出到本地，比如通过复制到剪贴板的方式导出，或者把它们直接保存成Access数据库文件或者Excel文件，或者利用“添加到已保存的数据源”导出成.tds或者.tdsx类型的数据源，还可以利用“数据提取”导出成.tde类型的数据源。如果想要与他人在网上共享数据源，可以将数据源发布到Tableau Online服务器或Tableau Server服务器上。

在Tableau Desktop中可以很容易把创建的工作表、仪表板和故事保存成图像或者打印为PDF，但这样制作出的分析报告将无法进行交互。

如果想要与他人共享工作簿，方法也有很多，可以把工作簿保存到本地，也可以发布到企业内部的Tableau Server服务器或者由Tableau托管的Tableau Online服务器上，还可以发布到完全免费和公开的Tableau Public服务器上。

第 11 章

Tableau Server简介

Tableau Server 是一种用于共享、分发和协作处理 Tableau 视图和仪表板的联机解决方案。用户通过Tableau Desktop完成视图设计后，可以将工作簿发布到Tableau Server，其他用户能够通过客户端（包括浏览器、移动终端、Tableau Desktop）看到工作簿内的所有视图并进行数据交互。Tableau Server为终端用户（End User）提供了进行简单修改的功能，但不支持创建新数据源、层级或计算字段。另外，Tableau Server也提供了集成服务，支持将视图通过多种方式嵌入到门户、网页、Web应用程序中。

11.1　安装 Tableau Server

根据企业数据架构、用户数、访问量的不同，Tableau Server可以灵活采用不同的配置方式。11.1.1节介绍最简单的单服务器安装过程，11.1.2节将以3节点集群为例介绍分布式集群安装。

11.1.1　单服务器安装

本节主要介绍在单台服务器上如何安装Tableau server。

1. 最低配置要求

安装之前要确保Tableau Server的计算机满足以下要求（见表11-1）。

- 操作系统：Tableau Server可以安装在Windows Server 2003（SP1或更高版本）、Windows Server 2008、Windows Server 2008 R2、Windows Vista 或Windows 7上。虽然Tableau Server可以在32 位操作系统上良好运行，但建议使用64 位操作系统。
- 内存、内核和磁盘空间：视服务器用户数而定。

表11-1　最低配置要求

部署类型	服务器用户数	CPU	RAM
评估	1、2	2内核	4 GB
小型	<25	4内核	8 GB
中型	<100	8内核	32 GB
企业	>100	16内核	>=32 GB

2. 运行安装程序

Tableau Server在Windows下的安装步骤如下。

(1) 双击安装文件，运行安装程序，见图11-1。

(2) 单击“Next”（下一步），按照屏幕指示完成安装。

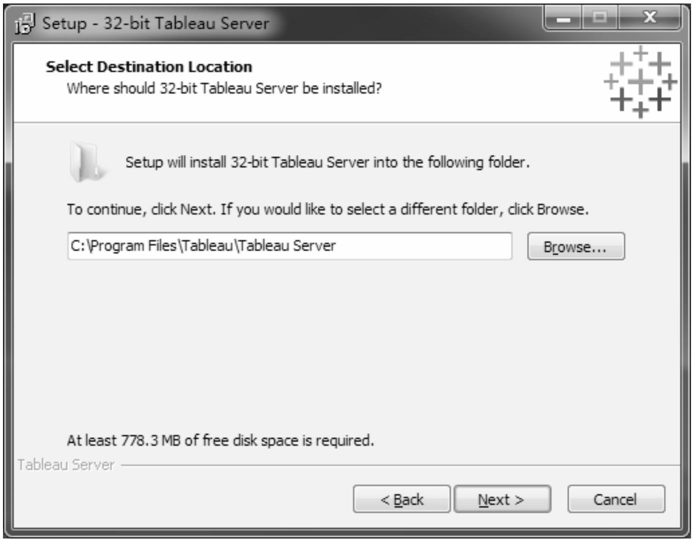


图11-1 安装应用程序

(3) 安装完成后，单击“Next”，打开“产品密钥管理器”窗口，见图11-2。

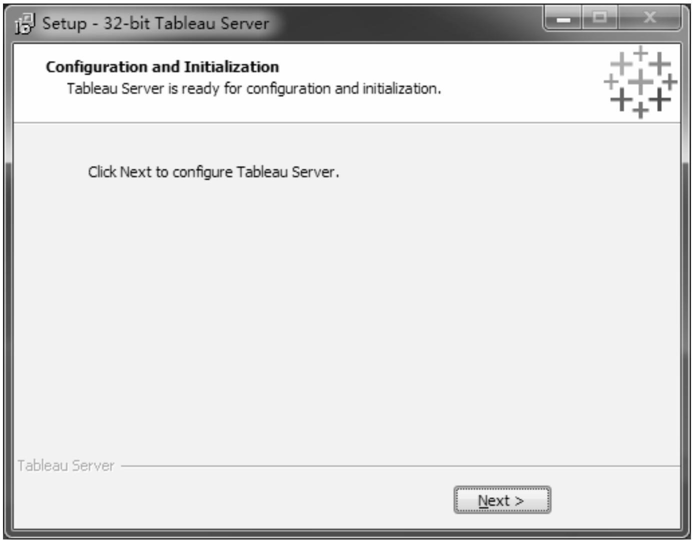


图11-2 安装完成

3. 产品激活

Tableau Server程序安装完成后进入产品激活界面。你可以选择“Start trial now”（现在开始试用），试用期为14天；也可以选择“Activate the product”（激活产品），输入产品密钥进行激活。

(1) 选择“Activate the product”，见图11-3。



图11-3　激活Tableau

(2) 将服务器产品密钥粘贴到相应的文本框中，然后单击“Activate”（激活）。如果计算机处于联网状态，Tableau将被顺利激活，见图11-4。

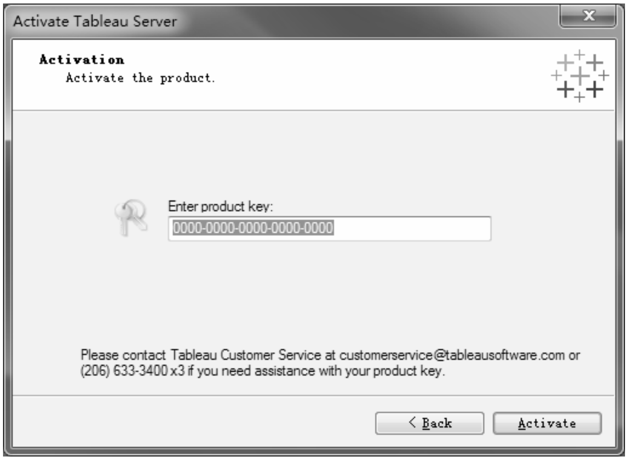


图11-4　在线激活Tableau

(3) 如果计算机处于脱机状态，激活将失败，系统会提供一个选项用于保存脱机激活文件。单击“Save”（保存），见图11-5。

11

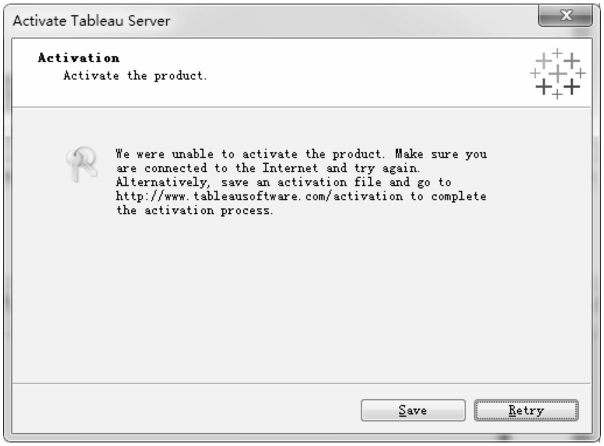


图11-5　保存脱机激活文件

(4) 为文件选择一个位置，然后单击“保存”，文件将保存为offline.tlq，见图11-6。

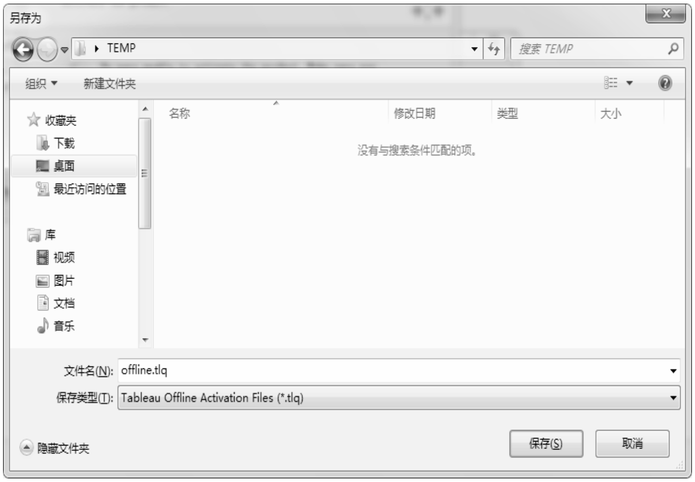


图11-6　选择脱机激活文件保存位置

(5) 返回Tableau，单击“退出”，关闭“激活”对话框。

(6) 从能够访问网络的计算机上打开Web浏览器并访问 Tableau 网站上的产品激活页，按照提示提交 offline.tlq 文件，提交完毕后页面会提示下载激活文件activation.tlf（参见图11-7）。

(7) 将activation.tlf文件复制到安装Tableau Server的计算机中。如果此计算机中已安装了Tableau Desktop，你可以双击这个新文件完成激活；如果没有安装Tableau Desktop，请继续执行步骤(8)和(9)。

(8) 在安装Tableau Server的计算机中，以管理员身份打开命令提示符并运行以下命令：`cd "C:\Program Files\Tableau\Tableau Server\\<version>\\bin"`，其中`<version>`是版本号。

(9) 在命令行输入tabadmin activate --tlf<path>\activation.tlf，其中<path>是activation.tlf所在的目录（例如tabadmin activate --tlf \Desktop\activation.tlf），将命令提示符窗口保持打开状态。

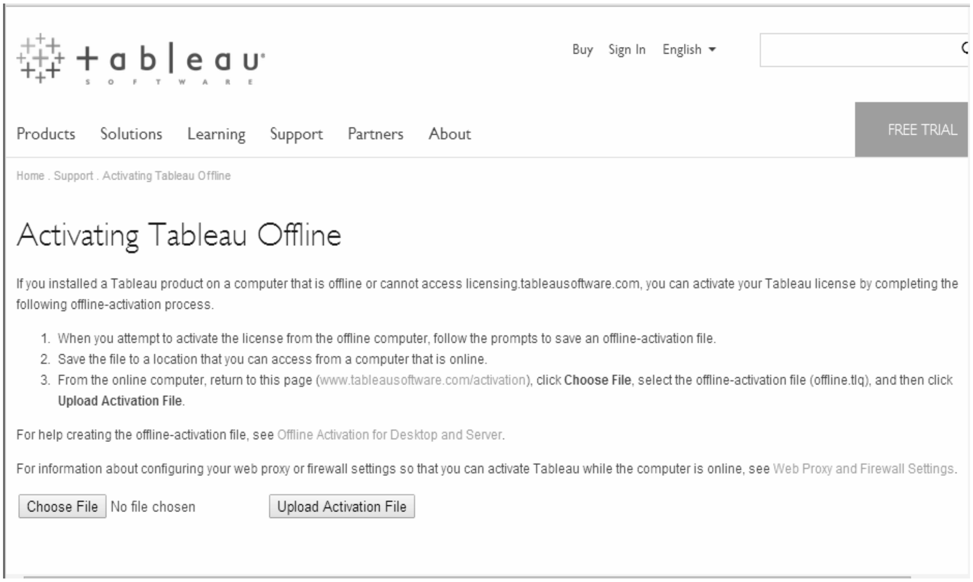


图11-7　Tableau在线激活

(10) 初始化许可证后将提示再次激活产品。在Tableau Server上，依次单击“开始”➤“所有程序”➤“Tableau Server 9.1”。

(11) 右键单击“管理产品密钥”，选择“以管理员身份运行”。即使作为管理员登录到Tableau Server计算机，也需要这样做以避免可能的注册错误。

(12) 单击“Activate the product”，再次输入产品密钥（与在步骤(1)中输入的产品密钥相同），保存.tlq文件。

(13) 从能够访问网络的计算机上打开 Web 浏览器，再次访问 Tableau 网站上的产品激活页，完成说明中的步骤。Tableau 将再次创建一个名为 activation.tlf 的文件并提示保存。

(14) 保存文件并将其复制到安装 Tableau Server 的计算机中。

(15) 返回到Tableau Server上的命令提示符窗口，键入tabadmin activate --tlf <path>\activation.tlf，其中<path> 是从产品激活页中保存的第二个响应文件的位置（例如tabadmin activate --tlf \Desktop\activation.tlf）。此时Tableau Server 已激活。

4. 完成基本配置

(1) Tableau Server默认在“网络服务”账户下运行。若要使用将为数据源提供 NT 身份验证功能的账户，请指定用户名和密码，用户名中应包含域名（参见图11-8）。

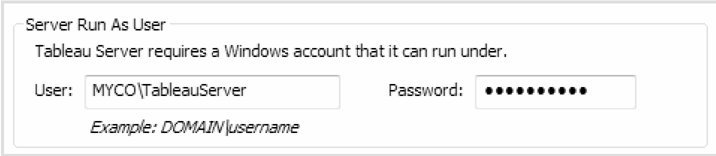


图11-8　指定用户名和密码

(2) 选择“域服务”（Active Directory）方式或“本地身份验证”（Local Authentication）方式对用户进行身份验证，见图11-9。如果选择本地身份验证方式，则会使用Tableau Server的内置用户管理系统来创建用户并分配密码；如果选择“域服务方式”，可以选择“Enable automatic logon”（启用自动登录），这会使用 Microsoft SSPI 基于用户的 Windows 用户名和密码自动登录，将创建类似于单点登录（SSO）的体验。如果计划将 Tableau Server 配置为使用 SAML、受信任身份验证或代理服务器，请勿选择“Enable automatic logon”。

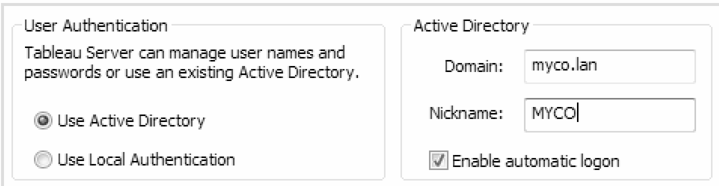


图11-9　选择用户身份验证方式

(3) 通过Web浏览器访问Tableau Server的默认端口为80，见图11-10。如果其他服务器正在端口 80 上运行或者有其他联网需要，则可能需要更改端口号。

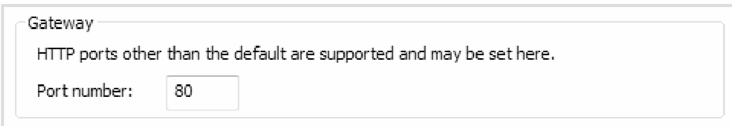


图11-10　Web访问Tableau Server端口设置

(4) 选择是否打开 Windows 防火墙中的端口。如果不打开此端口，则其他计算机上的用户可能无法访问该服务器，见图11-11。

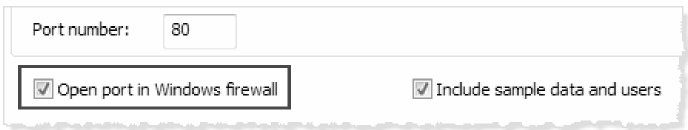


图11-11　打开Windows防火墙端口

(5) 选择是否包含示例数据和用户。“Include sample data and users”（包括示例数据和用户）

选项会安装若干示例工作簿和数据，可帮助熟悉 Tableau Server（特别是在安装该产品的试用版时）。如果选择"Include sample data users"，则会分配在 Tableau Server 中创建的第一个用户作为示例工作簿和数据的所有者。

(6) 可选择继续打开下一个页面以配置"缓存"和"初始 SQL"选项。如果不想配置这些选项，请单击"确定"。

5. 添加管理员账户

激活 Tableau Server 的最后一步是添加管理员账户，管理员将具有服务器的完全访问权限，包括管理用户、组和项目的权限。添加管理员账户的步骤根据使用域服务方式（Active Directory）或本地身份验证方式而有所不同。

- **Active Directory**

如果使用 Active Directory，请键入将成为管理员的现有 Active Directory 用户的用户名和密码，然后单击"添加用户"。注意：如果该管理员账户与服务器在同一个域中，则只需键入用户名，无需键入域；否则，应包括进完全限定域名，例如`test.lan\username`。

- **本地身份验证**

如果使用本地身份验证，请通过键入所选的用户名、显示名称和密码（键入两次）来创建管理员账户，然后单击"添加用户"。

11.1.2 分布式集群安装

通过分布式安装将Tableau Server的各个部分安装在不同的计算机上，这可以提高 Tableau Server 环境的可扩展性。

1. 分布式集群的部署模式

多节点的分布式集群通常包含一个主服务器节点、一个或多个工作服务器节点。典型部署模式包括：①3节点集群、②5节点集群、③高可用性集群、④虚拟机或基于云端部署。

分布式集群模式一方面可以利用更多的硬件提升服务器性能表现，一方面可以通过增加备用硬件提高可用性水平、缩短整个集群系统的可能停机时间。

Tableau最新的测试结果表明，在并发用户数为总用户数的10%、交互式用户占40%的访问负荷的假设下，4节点集群（4×16核CPU）能支持的总用户数达到5540，相比单节点服务器（16核CPU）所能支持的1900用户，节点数量的增加几乎实现了性能的同比例线性增长，见表11-2。

表11-2 分布式集群性能测试

节点数	并发用户数	总用户数
单节点	190	1900
2节点	270	2700
3节点	436	4360
4节点	554	5540

对于交互式用户占比为100%的访问负荷，测试结果见表11-3。

表11-3　分布式集群性能测试

节点数	并发用户数	总用户数
单节点	119	1190
2节点	206	2060
3节点	269	2690
4节点	347	3470

2. 集群部署安装

下面以3节点集群为例说明集群部署的安装方法。在本例中，3节点集群由1个主服务器和2个工作服务器构成。主服务器做网关，负责向2个工作服务器分配用户请求；2个工作服务器负责Tableau的核心进程，见图11-12。

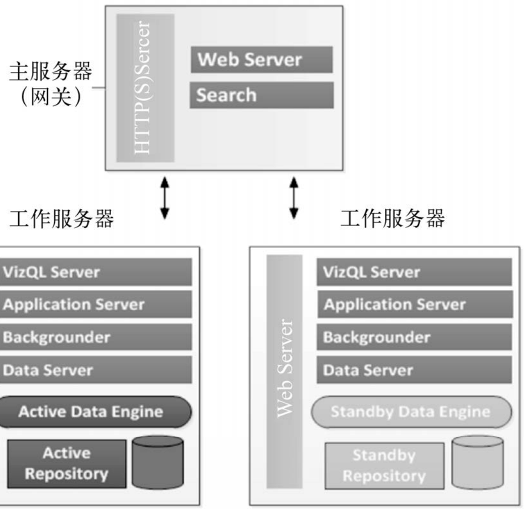


图11-12　节点服务器集群

(1) 应用服务（Application Server）：处理Web应用，支持浏览和搜索。
(2) 可视化（VizQL）：加载和呈现视图，计算和执行查询。
(3) 数据引擎（Data Engine）：存储数据提取和响应查询。
(4) 数据管理服务（Data Server）：管理Tableau服务器数据源连接。
(5) 后台（Backgrounder）：执行数据提取刷新任务、计划任务，以及通过tabcmd启动的任务。
(6) 存储库（Repository）：存储工作簿和用户元数据。

3. 节点集群安装配置过程

(1) 确保已在主计算机上安装 Tableau Server，停止主节点上的Tableau Server。

安装主服务器的方法在11.1.1节“单服务器安装”中已经进行了介绍。停止主服务器的方法为Windows“开始”➤“所有程序”➤“Tableau Server 8.0”➤“Tableau Server Monitor”➤“启动/停止服务器”。

(2) 在Tableau 客户账户中心下载Tableau Server工作软件，在要添加到Tableau Server群集的所有其他计算机上运行 Tableau Server 工作软件安装程序，安装过程中需提供主服务器的 IPv4 地址或计算机名称（建议使用计算机名称）。

(3) 添加工作服务器（Server Worker）。

保持主服务器停止状态，在主服务器的“配置Tableau Server”中选择“常规”选项卡并输入密码，选择“服务器”选项卡，然后单击“添加”，如图11-13所示。

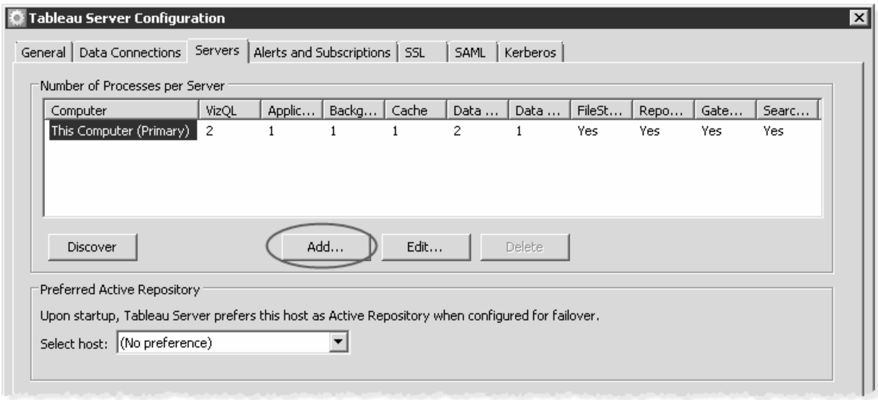


图11-13　分配工作服务器进程

在弹出的对话框中键入其中一台工作计算机的IPv4地址或计算机名称，并指定要分配给该计算机的“VizQL”“应用程序服务器”“后台程序”“缓存服务器”“数据服务器”“数据引擎”“文件存储”“存储库”“网关”和“搜索和浏览”进程的数量，见图11-14。在64位版本的工作服务器上，每一类进程最多可以添加两个实例，且每个服务器的最大总实例数为8。

添加完工作服务器后，单击“OK”以保存更改并关闭配置实用工具，需要几分钟时间才能完成更新。重复以上步骤，添加第二个工作服务器。

(4) 启动主节点上的 Tableau Server。

(5) 安装数据库驱动。

Tableau Server 和 Tableau Server Workers 的安装程序会自动安装 Oracle 和 Oracle Essbase 数据库的驱动程序。如果计划发布与其他数据库连接的工作簿和数据源，则需要确保主服务器和工作服务器上有相应驱动程序。

运行VizQL的工作计算机、应用程序服务器、数据服务器或后台程序进程需要这些数据库驱动程序。例如，如果有一台专门用作VizQL服务器的工作计算机和另一台专用于数据提取存储的计算机，则只需在VizQL服务器计算机中安装驱动程序。

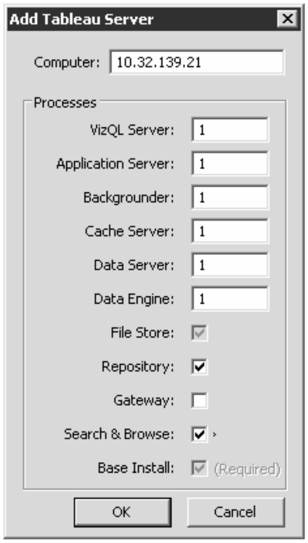


图11-14　分配工作服务器进程

完成上述步骤后，基本就完成了3节点集群的构建。

说明　完成集群设置后，可以在服务器的“Maintenance”（维护）页面上监视分布式计算机的状态，使用主服务器计算机上的命令行工具和配置工具进行配置更新，更新将被自动推送到工作计算机上。如果主服务器的IP 地址发生更改，则需要重新安装所有工作计算机。

11.2　配置 Tableau Server

本节主要介绍如何在Tableau Server中完成站点和用户权限的配置，从而完成系统初始化并实现业务信息的有效隔离。

11.2.1　配置站点

Tableau Server 允许服务器管理员为不同用户和内容集在服务器上创建多个站点，每个站点在服务器上是独立的，并且可以按用户或组设置项目、工作簿、视图或数据源的权限，每个站点的工作簿、数据和用户列表都独立于其他站点的相应项，并且用户一次只能访问一个站点。站点管理员（服务器管理员允许其创建站点用户）能够控制站点成员身份。服务器管理员创建站点之后，内容所有者可以将工作簿、视图和数据源发布到服务器上的特定站点。用户可以属于多个站点，在每个站点上具有不同的站点角色和权限。登录到服务器的用户将看到用户所属的站点中允许他们查看的内容。

服务器管理员可将站点添加到Tableau Server或编辑现有站点。即使在添加站点之前，Tableau Server也有默认站点。添加站点的步骤如下。

(1) 打开“Sites”（站点）页面。如果要在服务器上添加第一个站点，请选择“Settings”（设置）➤“Add a Site”（添加站点），然后单击“Add a Site”（添加站点）。否则，选择“Server”（服务器）➤“Site”（站点），然后单击“+New Site”（新建站点）。如图11-15所示。



图11-15　添加站点

(2) 为站点输入“Site Name”（站点名称）和“Site ID”（站点ID）（如果编辑的是Default站点，则无法更改“Site ID”），如图11-16所示。

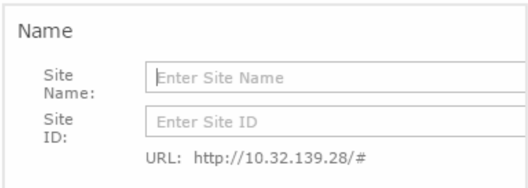


图11-16　输入站点名称和站点ID

(3) 工作簿、数据提取和数据源全部占用服务器上的存储空间。对于“Storage”（存储），选择“Server Limit”（服务器限制）或“GB”，并输入要作为限制的GB数（参见图11-17）。如果设置了服务器限制但站点超过了该限制，则将阻止发布者上传新内容，直至站点再次低于该限制。服务器管理员可以使用站点页面上的“Max Storage”（最大存储）和“Storage used”（已用存储）列来跟踪站点的限额使用情况。

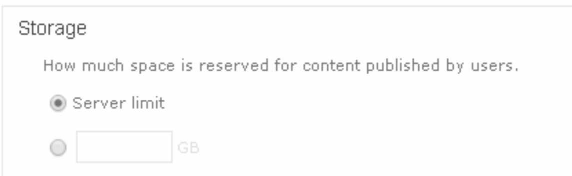


图11-17　存储空间设置

(4) 设置用户管理权限，Tableau Server可设为“Only server administrators”（只有服务器管理员可以添加和移除用户（仅服务器管理员）)，或“Server and site administrators”（服务器和站点管理员均可执行该操作），见图11-18。

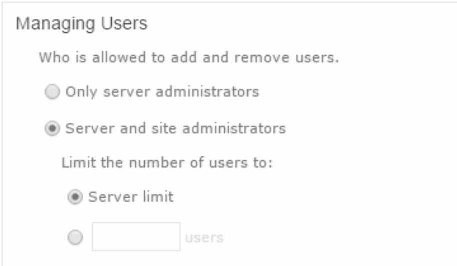


图11-18　设置管理用户权限

(5) 将“Allow users to use web authoring”（允许用户使用 Web 制作）保留为选中，禁用 Web 制作意味着用户无法从服务器 Web 环境编辑已发布工作簿。要更新已发布到服务器的工作簿，Tableau Desktop 用户必须重新发布它。

(6) 如果希望站点用户能够订阅视图，请保持“Allow users to subscribe to workbooks and views”（允许用户订阅工作簿和视图）为选中状态。只有当你也在“配置”对话框中配置了订阅设置时，此选项才可见。

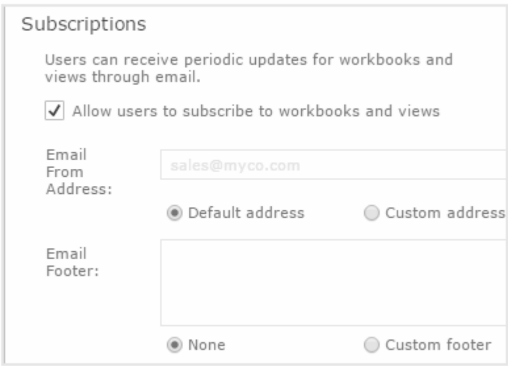


图11-19　订阅视图

(7) 选择“Allow to record workbooks performance”（允许记录工作簿性能指标），以允许站点用户收集有关工作簿性能的指标，例如工作簿加载速度。除了为站点选中此复选框，若要开始记录，用户还必须向工作簿的 URL 中添加参数。

(8) 单击“+New Site”或“Save”即创建站点成功。

若要编辑站点，请选择“Server”➤“Sites”（多站点），或单击“Settings”（单站点）。选择要修改的站点，然后选择“Edit Setting”（编辑设置）。

11.2.2 配置用户

需要访问 Tableau Server 的任何人（无论是发布、浏览还是管理内容的人）都必须作为用户添加。（对基于内核许可证的Tableau Server，我们可以创建“来宾”用户，即允许在服务器上没有账户的用户查看嵌入式视图并与其交互。）

服务器管理员可向服务器中添加用户，添加到 Tableau Server 的每个用户都必须具有关联的站点角色。站点角色由管理员分配，用于确定允许用户拥有的权限级别，包括用户是能够发布内容、与内容交互，还是只能查看发布到服务器的内容。关于如何添加用户并分配站点角色，请参见图11-20。

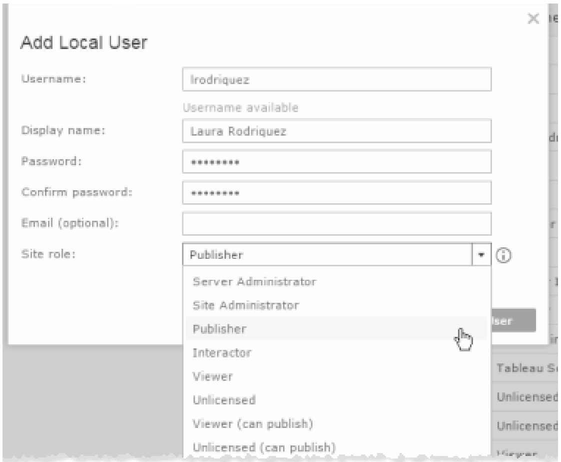


图11-20 添加用户并分配站点角色

(1) 服务器管理员（Server Administrator）：可访问服务器和所有站点上的所有服务器功能和设置。服务器管理员可创建站点，添加任何站点角色类型的用户，控制站点管理员是否能添加用户，创建其他服务器管理员，并且可以管理服务器自身，包括处理维护、设置、计划和搜索索引。

(2) 站点管理员（Site Administrator）：可以管理组、项目、工作簿和数据连接。默认情况下，站点管理员也可以添加用户以及分配站点角色和站点成员身份，此设置可由服务器管理员启用或禁用。站点管理员对特定站点上的内容具有不受限的访问权限。一个用户可被指定为多个站点上的站点管理员。

(3) 发布者（Publisher）：可以登录、浏览服务器，并且与已发布的视图交互，还可以从 Tableau Desktop 连接到 Tableau Server 以便发布和下载工作簿和数据源。

(4) 交互者（Interactor）：可以登录、浏览服务器，并且与已发布的视图交互，但不允许他们发布视图到服务器。

(5) 查看者（可发布）（Viewer (can publish)）：用户可从 Tableau Desktop 连接到 Tableau Server 以便发布和下载工作簿和数据源，但无法与服务器上的内容交互。

(6) 查看者（Viewer）：可以登录和查看服务器上的已发布视图，但无法与这些视图交互。

(7) 未许可（可发布）（Unlicensed（can publish））：无法登录到 Tableau Server，但可从 Tableau Desktop 连接到服务器以便将工作簿和数据源发布和下载到服务器。

(8) 未许可（Unlicensed）：无法登录到服务器。通过 CSV 文件导入服务器用户时，将为所有用户分配“未许可”站点角色。

Tableau有两种方式可以将用户添加到服务器或站点：①一次添加一个用户的方式，②批量导入的方式。此处以将用户添加到服务器为例介绍添加用户的操作步骤。

1. 一次添加一个用户

(1) 单击“Server“（服务器）➤ “Users”（用户），然后单击“+Add Users”（添加用户）（参见图11-21）。

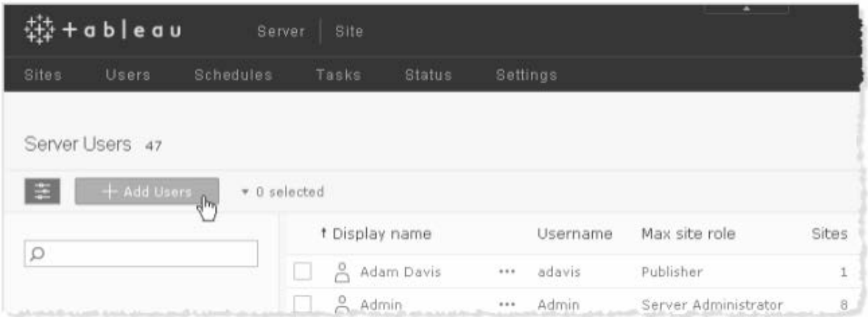


图11-21　一次添加一个用户

(2) 如果使用本地身份验证，请单击“Local user”（本地用户）。如果使用 Active Directory，请单击“Active Directory User”（Active Directory 用户），参见图11-22。

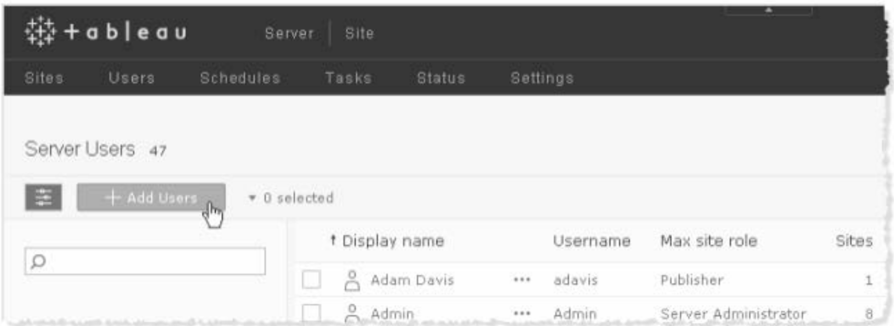


图11-22　一次添加一个用户

(3) 在对话框中，按照相应提示填写用户信息，包括用户名、显示名称、密码、确认密码、电子邮件以及站点角色，如图11-23所示。

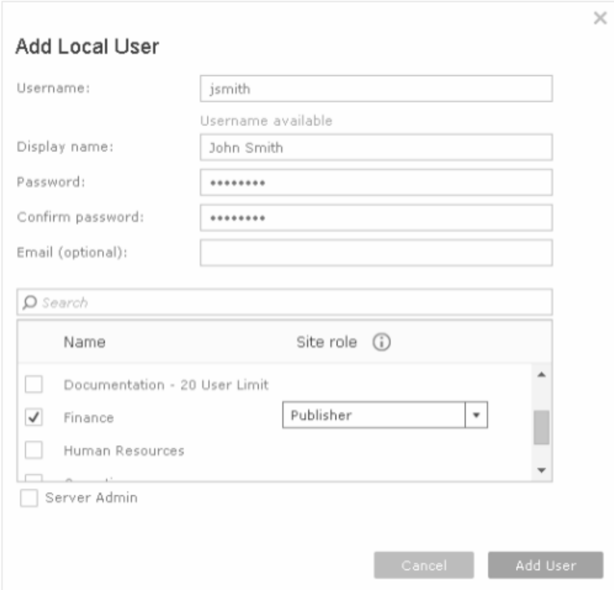


图11-23 填写用户信息

(4) 填写完成后，单击“Add User”（添加用户），即添加用户成功。

2. 批量导入用户

若要自动完成向服务器或站点添加用户的过程，我们需创建一个包含用户信息的 CSV 文件，然后导入该文件。

(1) 单击“Server”（服务器）➤“Users”（用户），然后单击“+Add Users”。

(2) 单击“Import From File（从文件导入），单击“Browse”（浏览）并导航到文件，然后单击“Import Users”（导入用户）（参见图11-24）。

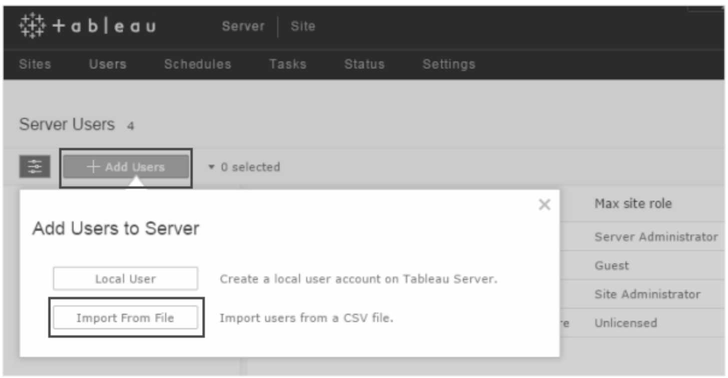


图11-24 从文件导入用户

11

(3) 在弹出的导入结果的对话框（参见图11-25）中单击“Done”（完成），实现批量导入。



图11-25　导入成功界面

11.2.3　配置组

简化用户管理的一种方法是将用户分配到组中。例如，可以向组分配权限，以便将权限应用到组中的所有用户。默认情况下每个站点中都存在“All Users”（所有用户）组。添加到服务器的每个用户都将自动成为“All Users”组的成员，我们无法删除此组，但可以为其设置权限。

组有两种创建方式：①本地创建、②通过 Active Directory 创建。

1. 创建本地组

本地组是使用 Tableau Server 内部用户管理系统创建的。操作步骤如下。

(1) 在站点中，单击“Groups”（组），然后单击“+New Group”（新建组）详见图11-26。

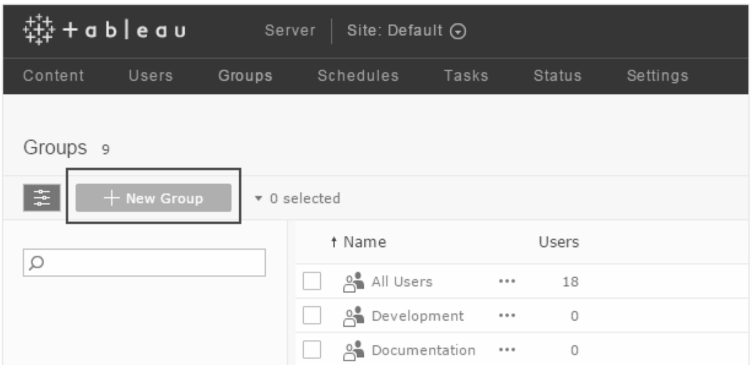


图11-26　新建组

(2) 为组键入一个名称，然后单击“Save”（保存），见图11-27。

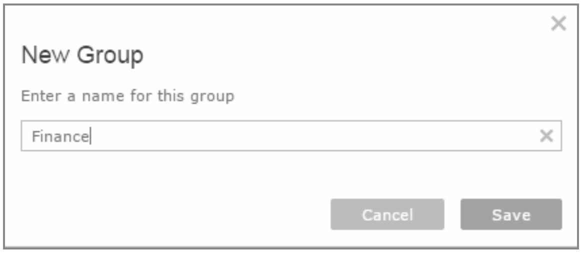


图11-27　键入组的名称

2. 通过Active Directory 创建组

通过导入 Active Directory 组，我们将在服务器上创建匹配组。导入步骤如下。

(1) 在站点中，单击“Groups”，然后单击“+Import Group”（导入组）。

(2) 键入要导入的 Active Directory 组的名称，然后在生成的列表中选择该组名称（见图11-28），并为组内用户选择站点角色。注意，无法在此处更改从 Active Directory 导入的组的名称，只能在 Active Directory 中更改。

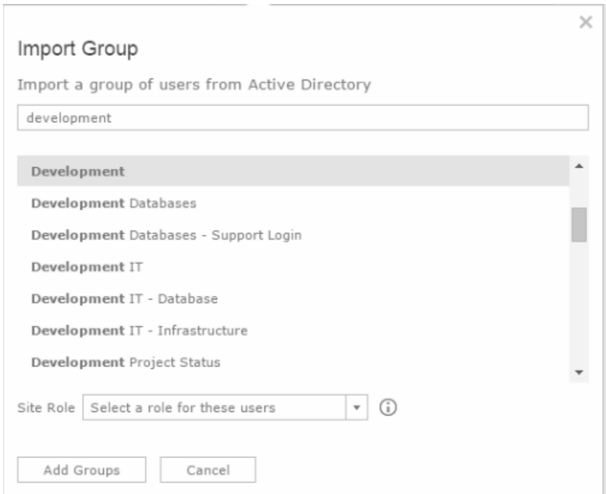


图11-28　键入组的名称

(3) 单击“Add Groups”（添加组），完成组的导入。

11.3　使用 Tableau Server

首次登录时，输入用户名和密码后单击“Login”（登录），从而登录到Tableau门户主页（参见图11-29）。



图11-29　门户主页

接下来我们介绍主页最常用的导航、两个工具和自定义用户首选项。

1. 导航

Tableau门户主页的左侧有“Projects”（项目）、“Workbooks”（工作簿）、“Views”（视图）和“Data Sources”（数据源）共4个导航标签。

- **项目**。是相关工作簿、视图和数据源的集合。只有管理员才可以创建新项目，并向用户和组分配“Project Leader”权限，此权限让用户能够指定项目权限以及将工作簿移到项目中。工作簿、视图或数据源可具有与项目权限不同的权限。例如，某个组可能没有查看项目 *X* 的权限，但可能有权查看发布到项目 *X* 的视图。

  工作簿的初始权限是从其项目权限复制而来的，视图的初始权限是从其工作簿权限复制而来的，但对项目的权限所做的更改不会自动应用于项目内的工作簿或工作簿内的视图，除非通过单击项目或工作簿的权限设置中的“Assign Permissions to Contents”（分配对内容的权限）特定分配新权限。
- **工作簿**。用户通过Tableau桌面版将工作簿发布到“Workbooks”，包括用户想要发布的仪表盘、工作表等。所有工作簿都必须位于项目中。默认情况下，工作簿将被添加到“Default“（默认）项目中。在创建自己的项目后，你可以将工作簿从一个项目移动到另一个项目中。
- **视图**。包含用户发布到Tableau Server上的所有仪表盘和工作表。
- **数据源**。数据源是一种可重用的数据连接，包含数据提取或与实时关系数据库直通连接有关的信息。

2. 工具

常用的工具有筛选器和收藏夹，如下所示。

- **筛选器**。可以非常简便地筛选内容。你可以利用“Search for content”（搜索框）输入想要搜索的内容，也可以利用“Filters”（筛选器）下的“Project”（项目）“Publisher”（发布者）“Modified on or after”（修改日期）等工具进行筛选或搜索。
- **收藏夹**（ ）。用于保存用户最喜欢的仪表盘。

3. 用户首选项

在“用户首选项”（User Preference）界面可以设置或修改用户的“Email Address”（邮箱地址）、“Start Page”（开始页）、“Language and Locale”（语言和地区）、“Account Password”（账户密码）等信息。

- **邮箱地址**。Tableau Server提供邮件订阅服务。这一服务允许用户选择想要关注的仪表盘，Tableau Server将根据系统管理员设置的时刻表按时向用户的邮箱地址推送信息。
- **开始页**。用户可以个性化地设置Tableau门户的开始页。开始页可以是项目层面的，也可以是某个特定的工作簿。
- **账户密码**。若服务器在安装时选择“Local Authentication”（本地身份验证）作为用户身份验证方式，则用户可以在该界面修改密码。

11.3.1 界面查询

进入仪表盘后，若权限允许，用户可以像在Tableau桌面版中一样与仪表盘交互，如可以使用筛选器、查看原始数据、下载仪表盘、订阅仪表盘、添加注释等。



图11-30 在线交互

1. 使用筛选器

在仪表盘中上端的筛选器中可以自由筛选，如初始设置为“2014年5月”，改选“2014年6月”后仪表盘发生变化。

2. 查看原始数据

选择某一标记区块，在弹出框单击“View Data”（查看数据）。在新对话框中，单击“Underlying”➤“Show all columns”查看所有列。

3. 订阅

单击☒邮件图标，填写“Subject”（主题）为邮件标题（默认为仪表盘的名称），单击“Schedule”（计划安排）下拉框以选择推送时间，选择“Content”（内容）以选择推送内容为“The Sheet”（这一工作表）或“Sheets in Workbook”（工作簿中所有工作表），检查电子邮件地址是否正确（若有误可至“用户首选项”界面修改，详见11.2.2节）。完成后，单击“Subscribe”（订阅），则订阅内容会以静态图片形式按时推送到你设置的邮箱中；单击邮件正文中的图片将导向Tableau Server门户网站上的相应仪表盘。

4. 添加评论

仪表盘底部为评论区。在“Comments”（评论）文本框中输入内容，单击“Add”（添加），则其他用户可以看到你对该仪表盘的评论，如图11-31所示。



图11-31　添加评论

5. 保存改动

单击“Remember my Changes”（记住我的改动），输入新视图名称（如“重庆-6月”），单击“Remember”（记住），如图11-32所示。通过在服务器上保存视图，用户可以与其他用户分享自己的视角。



图11-32 记住我的选择

这一动作成功后将有特定的URL地址指向这一视图，如单击“重庆-6月”指向这一视图。单击“Manage Custom Views”（管理自定义视图）来进入自定义视图管理界面（见图11-33）。



图11-33 自定义视图管理

若需要分享某一视图，请单击该视图名称旁的图标，然后单击“Done”（完成）保存改动，见图11-34。

11

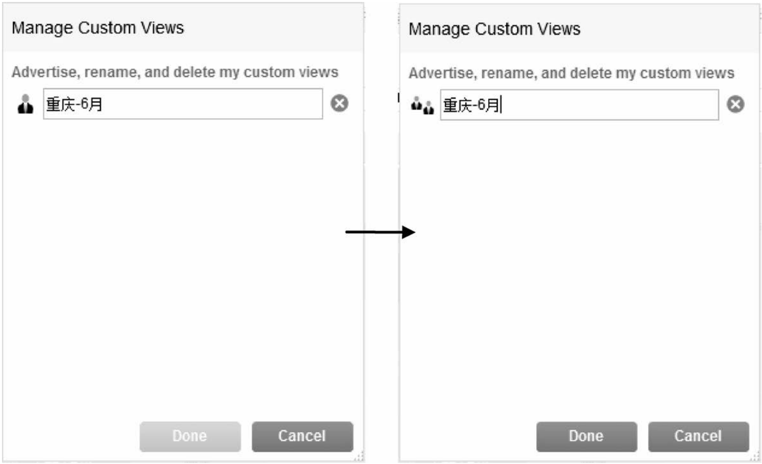


图11-34　共享自定义视图

6. 4个功能键

保存后回到仪表盘，在仪表盘顶端可看到4个按键（如图11-35所示），从左至右依次为“内容导出”“全部恢复”“暂停”和“刷新”按键。



图11-35　将仪表盘内容导出到本地（左一）

- **“内容导出”**：支持用户将仪表盘以图片或PDF格式导出到本地计算机，或者以数据或交叉表形式将信息导出到本机。
- **“全部恢复”**：允许用户清除交互，将仪表盘恢复到初始状态。
- **“暂停”**：允许用户暂停仪表盘的自动更新。当用户准备对仪表盘做一系列交互动作时，可以按下“暂停”键，则服务器暂停向数据源自动更新数据，直到用户再次按下“暂停”键。
- **“刷新”**：允许用户手动更新数据。

11.3.2　编辑发布

通过单击左上角的“Edit”（编辑），用户可以对视图、仪表盘进行简单的在线编辑操作。

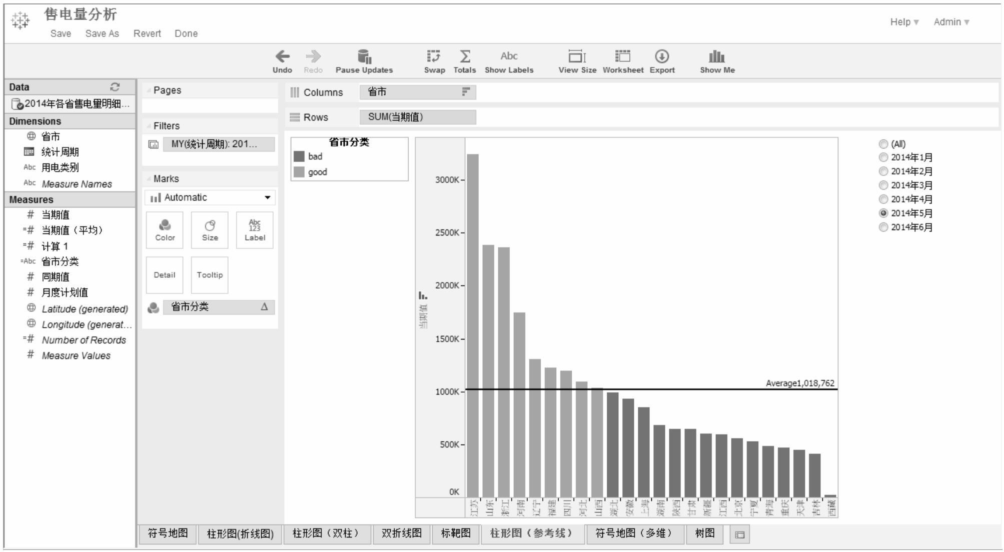


图11-36　编辑

Tableau Server的在线编辑功能与Tableau桌面版非常相似，但具有局限性：用户不能创建新的数据源、新的层级（hierarchy）、新的计算字段（calculate field）。它只支持用户通过最基本的拖放动作进行简单的编辑、创建新的工作表等操作。

若用户有“发布”权限，则可以使用“Save As”（另存为）将新工作表保存到本工作簿，如图11-37所示。你也可以使用“Save”（保存）保存其他任何修改过的工作表。

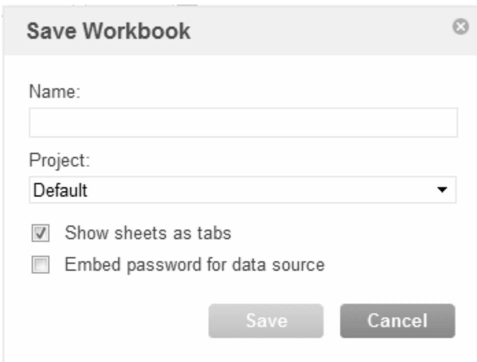


图11-37　另存工作簿

11.4　安全机制

Tableau Server主要包含4类安全机制：访问安全、对象安全、数据安全、网络安全。

11.4.1　访问安全

Tableau Server 支持3类身份验证：①Active Directory，针对Tableau Server所使用的Windows用户进行验证。②本地身份验证，使用Tableau Server 的内部身份验证机制。③受信任的身份验证，基于Tableau Server与一个或多个Web 服务器之间的信任关系进行身份验证。

1. Active Directory

使用Active Directory 进行身份验证时，所有用户名和密码都由Active Directory 进行管理。当用户在Tableau Server登录时输入凭据时，Tableau 会将凭据传递给Active Directory 服务器。

通过Active Directory 进行用户身份验证意味着将使用本地计算机的Windows凭据使用用户自动登录，而不是从可能已登录的其他系统或门户传递凭据。例如，如果用户以“MSmith”身份登录其本地计算机，然后以“Mary”身份登录到一个SharePoint 门户，则访问Tableau Server使用的凭据将是“MSmith”的凭据。若要使用来自SharePoint 站点的凭据(“Mary”)自动登录，SharePoint门户必须使用采用受信任的身份验证方式的Tableau Web组件。

2. 本地验证

通过本地身份验证进行用户身份验证时，Tableau Server管理用户、组、密码以及整个身份验证过程。管理员可方便地将用户列表导入到Tableau Server，并可通过tabcmd方式执行大多数用户管理功能。用户可在看到提示时输入凭据来完成手动登录，或在访问门户中的内容时通过透明的受信任的身份验证来登录。

3. 受信任的身份验证

受信任的身份验证意味着你已在Tableau Server与一个或多个Web 服务器之间建立受信任的关系。例如，采用受信任的身份验证后，登录到门户的员工可以直接查看仪表板，而无需其他登录。

当Tableau Server接收来自受信任Web 服务器的请求时，它会假设该Web 服务器已处理必要的身份验证。Tableau Server 接收带有可兑现令牌或票证的请求，并向用户显示考虑进用户的角色与权限的个性化视图。

11.4.2　对象安全

在Tableau Server中，对象安全主要包含以下3个概念。

- **权限**：决定了用户可以访问的对象、内容，以及可以执行的操作。
- **角色**：包含一组权限，具有该角色的用户将拥有这组权限。
- **许可级别**：用户可拥有的最大权限集。

在Tableau中，一个“角色”就是一组加到内容上的“权限”，它控制用户与项目、工作簿等对象的交互能力。指派给用户的“角色”针对特定视图，而非针对系统中的所有内容。举例来说，一个用户针对某一特定视图被指派“交互者”的角色，但该用户针对另一视图的权限也许只是“浏

览者”。许可级别控制一个用户的最大权限集。比如，一个具有“浏览者”许可级别的用户，不管针对某一视图被赋予什么样的角色，他都无法使用筛选器（因为使用筛选器需要“交互”权限）。

基于“角色”与“许可级别”，对象安全用于控制用户能够访问的对象和可以执行的操作，但不控制视图中显示哪些数据。用户能否访问数据由数据安全机制控制。

11.4.3 数据安全

Tableau为控制哪些用户可以查看哪些数据提供了3种方法：数据库验证、Tableau安全验证以及混合验证。

- **数据库登录账户**。创建连接到实时数据库的数据源时，可以选择是通过 Windows NT 还是数据库的内置安全机制针对数据库进行身份验证。
- **身份验证模式**。发布具有实时数据库连接的数据源或工作簿时，可以选择一种“身份验证模式”；可用的模式取决于上面的选择。
- **用户筛选器**。可以在工作簿或数据源中设置筛选器，以基于用户的 Tableau Server登录账户控制其可在已发布视图中查看的数据数据库验证。

表11-4概括了上述选项的优势和劣势。

表11-4 数据库验证

数据库连接选项		数据安全问题		
数据库登录账户	身份验证模式	是否每个Tableau Server用户都可实现数据库安全性	用户筛选器是否为限制各用户可查看的数据的唯一方法	Web缓存是否在用户间共享
Window NT集成安全性（Windows身份验证）	用户运行身份账户	否	是	是
	通过服务器运行身份账户进行模拟	是	否	否
用户名和密码	提示用户：当查看者查看视图时，提示其提供数据库凭据。可以保存凭据	是	否	否
	嵌入式凭据：工作簿或数据源发布者可以嵌入其数据库凭据	否	是	是
	通过嵌入式密码进行模拟：嵌入带IMPERSONATE权限的数据库凭据	是	否	否

11.4.4 网络安全

Tableau Server主要包含3类网络传输接口。①客户端到Tableau Server：客户端可以是Web浏览器、Tableau Desktop 或tabcmd实用工具。②Tableau Server 到数据库：TableauServer 与数据库间进行通信，以刷新提取数据或处理实时数据库连接。③服务器组件通信：适用于分布式部署下各服务器间通信。

1. 客户端到Tableau Server

Tableau Server 客户端可以是Web 浏览器、Tableau Desktop 或tabcmd。在内网环境下，Tableau Server与其客户端之间的通信可以使用标准HTTP 请求和响应；在对安全性要求较高的环境下，也可以采用HTTPS（SSL）协议。当采用SSL方式时，客户端和服务器之间的所有内容和通信都将被加密。

客户端和服务器之间的密码通信使用公钥/私钥加密。Tableau Server向客户端发送一个公钥，客户端使用此密钥对密码加密进行传输。每次加密传输都使用一次性密钥，传输完后更换新的密钥。这意味着无论是否使用SSL，密码都始终是安全的。

2. Tableau Server 到数据库

Tableau Server动态连接到数据库以获取结果集并刷新提取数据时，服务器会优先使用本机驱动程序新建数据源库，在没有本机驱动程序时会使用通用的ODBC驱动。本机驱动程序安装过程中将完成配置驱动程序以在非标准端口上通信或提供传输加密，这种配置类型对Tableau 是透明的。

3. 服务器组件通信

分布式服务器安装中的Tableau Server 组件间通信有两个方面：信任和传输。Tableau群集中的每个服务器都使用一个严格信任模型来确保其接收来自群集中其他服务器的有效请求。主服务器是群集中唯一接受第三方（客户端）请求的计算机，群集中的所有其他计算机都只接受来自群集中其他受信任成员的请求。信任是通过IP 地址、端口和协议的白名单建立的，如果任何一部分无效，请求都会被忽略。群集的所有成员都可以互相通信。除许可证验证和访问存储库外，所有内部通信的传输都是通过HTTP执行的。

附录 A

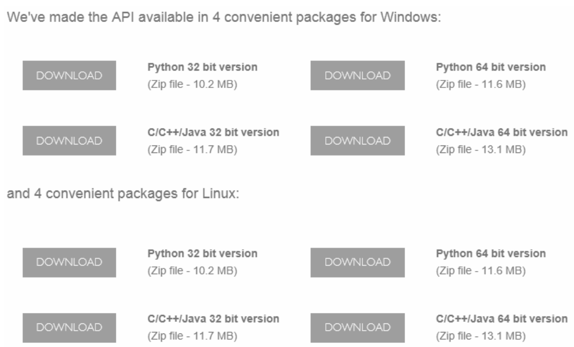
Tableau数据提取API

Tableau可以连接使用多种格式的数据源，如Microsoft execl、csv、txt以及多种数据库。但由于业务的复杂性，数据源的格式多种多样，Tableau所支持的数据源格式仍不可能面面俱到。那么，Tableau用户遇到Tableau不识别的数据源时该怎么办？

Tableau提供了数据提取API来解决这个问题。在遇到Tableau不识别的数据源格式时，Tableau利用数据提取API将它转化为Tableau的数据提取（tde格式文件）。数据提取API参见Tableau网站：http://www.tableausoftware.com/data-extract-api。

A.1　获取数据提取API版本及示例运行

Tableau提供了图A-1所示的8种数据提取API，支持Windows平台和Linux平台，在每个开发包中都有使用示例，开发人员可以选择相应的版本进行下载。Tableau数据提取API目前支持4种开发语言，本节将介绍4种开发语言如何使用数据提取API，而示例代码的功能是创建一个数据提取。本节以Windows 32位API版本为例进行描述。



图A-1　数据提取API版本

A.2　C/C++运行示例

本示例中C/C++使用数据提取API运用到了C++ 11中的特性，所以需要C++ 11的编译器，其

使用步骤如下。

(1) 将下载的TDE-API-C-Java-32Bit.zip解压缩，${ExtractDirectory}是解压缩后的路径。

(2) 将${ExtractDirectory}\bin添加到系统的%PATH%环境变量中，这个目录存放的是数据提取API使用的动态链接库。

(3) 在visual stdio中新建项目，新建控制台项目。

说明 数据提取API要求visual stdio的版本不低于2008，所以请使用visual stdio 2010及以后的版本。

---

(4) 将${ExtractDirectory}\samples下的MakeOrder.c或MakeOrder.cpp添加到新建的项目中。

(5) 选择菜单栏中“项目”➤“属性”，在弹出的属性页中选择“C/C++”➤“常规”，将数据提取API的头文件目录${ExtractDirectory}\include添加到“附加包含目录”。

(6) 选择菜单栏中的“项目”➤“属性”，在弹出的属性页中选择“连接器”➤“常规”，将数据提取API编译所需的lib文件目录${ExtractDirectory}\lib添加到“附加库目录”。

(7) 选择菜单栏中“项目”➤“属性”，在弹出的属性页中选择“连接器”➤“输入”，将数据提取API编译所需的lib文件DataExtract.lib添加到“附加依赖项”。

(8) 编译并运行。

A.3 Java运行示例

本示例中Java使用数据提取API的步骤如下。

(1) 将${ExtractDirectory}\bin添加到系统的%PATH%环境变量中，这个目录存放的是数据提取API使用的动态链接库。我们在Java中使用了这些动态链接库。

(2) 在Eclipse中新建Java项目。

(3) 将${ExtractDirectory}\samples目录下的com文件夹复制到新建工程的SRC下。

(4) ${ExtractDirectory}\java6下的dataextract.jar和jna.jar作为外部JAR添加到新建项目的生成路径下。

(5) 编译并运行。

说明 在${ExtractDirectory}目录中还有一个Java文件夹，里面也有dataextract.jar和jna.jar，与Java 6的区别在于，Java文件中的jar使用JDK 1.7编译的，而Java 6中使用JDK 1.6编译的。

A.4 Python运行示例

本示例中Python使用数据提取API的步骤如下。

(1) 将下载的TDE-API-Python-32Bit.zip解压缩，${ExtractDirectory}是解压缩后的路径。

(2) 在命令行中，进入${ExtractDirectory}\DataExtract-8200.14.0720.2105目录，这个目录名称随着Tableau发布不同版本API会在日期部分稍有不同。执行“python setup.py install”命令安装数据提取API的模块。

(3) 在命令行中进入${ExtractDirectory}\DataExtract-8200.14.0720.2105\samples目录，执行python makeorder.py命令。

A.5　数据提取API使用介绍

上面主要介绍了下载API后的创建及编译运行过程，本节将简要阐述API的基本过程。目前提供的4种开发语言尽管在语法上存在差异，但是原理基本一致，为以下5个基本步骤。

(1) 创建Extract对象；创建该对象会生成一个TDE文件。

(2) 给新创建的数据提取添加表。

(3) 在表中添加字段。

(4) 根据字段，向表中插入数据。

(5) 释放资源。注意，资源释放对编写来说很重要，所以一定要注意资源的释放。

这里我们整理了常用语言Java的API列表（与C++的类情况非常类似），以方便读者查阅，参见表A-1。

表A-1　Java数据提取API

类（Class）	描述（Description）
TableDefinition	用于定义数据提取对象中各个列的字段名称与字段类型信息（首列为0） 构造函数： ❑ TableDefinition()：创建空的TableDefinition对象 方法： ❑ void addColumn(String name, Type type)：向TableDefinition添加新列的字段名称和字段类型 ❑ void addColumnWithCollation(String name, Type type, Collation collation)：向TableDefinition添加新列的字段名称、字段类型以及编码整理 ❑ Collation getColumnCollation(int columnNumber)：返回特定列的编码整理方式 ❑ int getColumnCount()：返回TableDefinition定义的列数量 ❑ String getColumnName(int columnNumber)：返回特定列的字段名称 ❑ Type getColumnType(int columnNumber)：返回特定列的字段类型
TableDefinition	❑ Collation getDefaultCollation()：返回默认的编码整理方式（若没有设置，默认为binary） ❑ void setDefaultCollation(Collation collation)：为新加的字符类型字段设置编码整理方式 ❑ Pointer getHandle()：返回TableDefinition实例的句柄 ❑ void close()：关闭TableDefinition，并释放内存空间 ❑ WString TabGetLastErrorMessage()：返回错误的描述信息

（续）

类（Class）	描述（Description）
Table	用于创建Tableau数据提取中的数据表（Table） **构造函数：** ❑ 注意，请使用Extract类中的addTable方法创建数据表对象 **方法：** ❑ void insert(Row row)：向数据表（Table）中插入行记录 ❑ TableDefinition getTableDefinition()：返回数据表（Table）的TableDefinition ❑ Pointer getHandle()：返回Table实例的句柄 ❑ WString TabGetLastErrorMessage()：返回错误的描述信息
Row	用于插入到数据提取中数据表的行记录数据 **构造函数：** ❑ Row(TableDefinition tableDefinition)：创建具有特定TableDefinition的空行对象 **方法：** ❑ void setBoolean(int columnNumber, boolean value)：设定该行数据对象中特定列字段的布尔值 ❑ void setCharString(int columnNumber, String value)：设定该行数据对象中特定列字段的字符串值 ❑ void setString(int columnNumber, String value)：设定该行数据对象中特定列字段的字符串值 ❑ void setDate(int columnNumber, int year, int month, int day)：设定该行数据对象中特定列字段的日期值（年、月、日） ❑ void setDateTime(int columnNumber, int year, int month, int day, int hour, int min, int sec, int frac)：设定该行数据对象中特定列字段的日期/时间值（年、月、日、时、分、秒、毫秒） ❑ void setDuration(int columnNumber, int day, int hour, int minute, int second, int frac)：设定该行数据对象中特定列字段的持续时间值（日、时、分、秒、毫秒） ❑ void setDouble(int columnNumber, double value)：设定该行数据对象中特定列字段的双精度浮点值 ❑ void setInteger(int columnNumber, int value)：设定该行数据对象中特定列字段的整型值 ❑ void setLongInteger(int columnNumber, long value)：设定该行数据对象中特定列字段的长整型值 ❑ void setNull(int columnNumber)：设定该行数据对象中特定列字段为空值 ❑ Pointer getHandle()：返回行记录数据（Row）实例的句柄 ❑ void close()：关闭行记录数据对象（Row），并释放内存空间 ❑ WString TabGetLastErrorMessage()：返回错误的描述信息
Extract	用于创建Tableau数据提取对象，并向其中添加表 **构造函数：** ❑ Extract(String path)：创建数据提取文件对象，path为创建.tde文件的路径 **方法：** ❑ Table addTable(String name, TableDefinition tableDefinition)：向数据提取添加表，并返回该新添加表的指针

（续）

类（Class）	描述（Description）
Extract	❑ boolean hasTable(String name)：检测数据提取中是否存在该表 ❑ Table openTable (String name)：打开数据提取中存在的表，并返回其指针 ❑ Pointer getHandle()：返回Extract实例的句柄 ❑ void close()：关闭数据提取，并关闭打开的全部表 ❑ WString TabGetLastErrorMessage()：返回错误的描述信息
Collation（枚举类）	枚举全部支持的字符串编码整理方式 **枚举常量**：共有43个成员，其中常用的有ZH_HANS_CN、BINARY等 **方法**： ❑ Collation enumForValue(int value)：返回给定值的编码整理方式 ❑ Collation valueOf(String name)：返回给定名称的编码整理方式 ❑ Collation[] values()：返回含有编码整理方式的数组 ❑ int getValue()：返回特定实例的编码整理方式值
Type（枚举类）	枚举全部支持的数据类型 **枚举常量**：共有8个成员，包括BOOLEAN、CHAR_STRING、DATE、DATETIME、DOUBLE、DURATION、INTEGER、UNICODE_STRING。其中，中文时使用UNICODE_STRING。 **方法**： ❑ Type enumForValue(int value)：返回给定值的数据类型 ❑ Type valueOf(String name)：返回给定名称的数据类型 ❑ Type [] values()：返回含有数据类型的数组 ❑ int getValue()：返回特定实例的数据类型值
Result（枚举类）	枚举全部返回值 **枚举常量**：共有23个成员，其中常见的有SUCCESS、WRONG_TYPE、OUT_OF_MEMORY、UNKNOWN_ERROR等 **方法**： ❑ Result enumForValue(int value)：返回给定值的编码整理方式 ❑ Result valueOf(String name)：返回给定名称的编码整理方式 ❑ Result [] values()：返回含有编码整理方式的数组 ❑ int getValue()：返回特定实例的编码整理方式值
TableauException	用于抛出Tableau代码中的异常 **构造函数**： ❑ TableauException(int errorCode, String message)：创建Tableau抛出异常的对象 **方法**： ❑ int getErrorCode()：返回异常抛出的错误代码

附录 B

Tableau JavaScript API

当用户想访问发布到服务器上的视图时，可以通过Web方式连接Tableau 服务器，但这种情况需要用户有专有的Tableau服务器账户，这对于其他外部应用访问或只是简单浏览工作簿的用户来说是不实际的。那么如何在一个Web应用上让用户访问Tableau 工作表和仪表板，而不是让用户简单地直接连接Tableau服务器？Tableau JavaScript API可以使Web应用嵌入工作表和仪表板以进行访问交互，而不是直接连接Tableau服务器。接下来我们将介绍如何在Web应用中使用Tableau JavaScript API。本附录只是简单介绍Tableau JavaScript API的使用准备和简单原理，若要详细了解API的使用请参考Tableau Server的联机帮助文档：http://onlinehelp.tableausoftware.com/current/server/zh-cn/help.htm。

B.1 使用要求

使用Tableau JavaScript API有以下两点要求。

- 能够访问Tableau服务器。使用Tableau JavaScript API需要可以访问Tableau Server、Tableau Online和Tableau Public中的任意一个，包括发布到其中的工作簿。
- 受支持的浏览器。最终用户可以在大多数受支持的 Web 浏览器中使用你创建的 Web 应用程序，具体而言就是：Chrome、Firefox、Safari 3.2.1 和更高版本，以及 Internet Explorer 8.0 和更高版本。如果你使用的是 Internet Explorer 8.0，则必须禁用兼容性模式。另外，浏览器必须配置为允许在 JavaScript 代码中使用 `window.postMessage` 方法。一些安全包会禁用此功能。

B.2 访问JavaScript API

该 API 通过文件 tableau_v8.js（缩小）或 tableau_v8.debug.js 提供。在包含用于呈现 Tableau 视图的 JavaScript 代码的网页中，添加以下代码：

```
<script type="text/javascript" src="http://localhost/javascripts/api/tableau_v8.js"> </script>
```

其中`src`是`js`代码所在的服务器地址，根据所访问Tableau服务器的不同可以使用以下地址引用 API 文件。

(1) 对于Tableau Server，使用以下位置：

```
src="http://<your_server_name>/javascripts/api/tableau_v8.js"
```

(2) 对于 Tableau Online，使用以下位置：

```
src="https://online.tableausoftware.com/javascripts/api/ tableau_v8.js"
```

(3) 对于 Tableau Public，使用以下位置：

```
src=https://public.tableausoftware.com/javascripts/api/ tableau_v8.js
```

说明　为获得最佳的代码稳定性，请使用所用服务器产品的 API 位置。例如，如果你将工作簿发布到本地 Tableau Server，请使用服务器的本地 JavaScript API。

B.3　连接到服务器模式

根据是否需要身份验证，通过JavaScript连接Tableau 服务器有两种方式。以下是用两种方式进行连接的代码示例。

1. 不需要身份验证

```
var placeholderDiv = document.getElementById("tableauViz");
var url = "http://tabserver/views/workbookname/viewname";
var options = {
   hideTabs: true,
   width: "800px",
   height: "700px"
};
var viz = new tableauSoftware.Viz(placeholderDiv, url, options);
```

其中placeholderDiv是在HTML中显示Tableau视图的容器，url是视图所在的位置，options是视图显示时的相关属性。

2. 需要身份验证

```
var placeholderDiv = document.getElementById("tableauViz");
var url = "http://tabserver/trusted/Etdpsm_Ew6rJY-9kRrALjauU/views/workbookname/viewname";
var options = {
   hideTabs: true,
   width: "800px",
   height: "700px"
};
var viz = new tableauSoftware.Viz(placeholderDiv, url, options);
```

不需要身份验证的地方已用粗体显示，其中/trusted表示Tableau服务器需要身份验证，/Etdpsm_Ew6rJY-9kRrALjauU是身份验证字符串。验证字符串的获取需要经过以下步骤。

(1) Web应用所架设的服务器被连接到的Tableau服务器信任。该信任关系在Tableau服务器端

进行设置。

(2) 用户通过浏览器发送GET请求给Web应用。

(3) Web应用发送带有账号名、客户端IP等信息的POST请求给Tableau服务器。

(4) Tableau服务器将对相关信息进行验证，生成验证字符串，并返回给Web应用。

(5) Web应用接收到验证字符串后，将带有验证字符串的HTML代码发送给客户端浏览器。

(6) 客户端浏览器通过带身份验证字符串的URL获取视图。

附录 C

Tableau函数

C.1　数字函数

序　号	函　数	含　义	示　例
1	ABS	返回给定的绝对值	ABS(-7)=7
2	ACOS	返回数字的反余弦，结果以弧度表示	ACOS(-1)=3.14159265358979
3	ASIN	返回数字的反正弦，结果以弧度表示	ASIN(1)=1.5707963267949
4	ATAN	返回数字的反正切，结果以弧度表示	ATAN(180)=1.5652408283942
5	ATAN2	返回两个给定数(*x*和*y*)的反正切，结果以弧度表示	ATAN2(2,1)=1.10714871779409
6	COS	返回角度的余弦，以弧度为单位指定角度	COS(PI()/4)=0.707106781186548
7	COT	返回角度的余切，以弧度为单位指定角度	COT(PI()/4)=1
8	DEGREES	将以弧度表示的数字转换为度数	DEGREES(PI()/4)=45.0
9	DIV	返回除法的整数部分	DIV(11,2)=>5
10	EXP	返回e的给定数字次幂	EXP(5)=e^5
11	LN	返回给定数字的自然对数；如果数字小于或等于0，则返回Null	\
12	LOG	返回数字以给定底数为底的对数。如果省略了底数值，则默认底数值为10	\
13	MAX	返回单个表达式在所有记录间的最大值，或两个表达式对于每个记录的最大值	MAX([Sales])
14	MIN	返回一个表达式在所有记录间的最小值，或两个表达式对于每个记录的最小值	MIN([Profit])
15	PI	返回数字常量pi	\
16	POWER	返回数字的给定幂的结果	POWER(5,2)=5^2=25

（续）

序　号	函　数	含　义	示　例
17	RADIANS	将以度数表示的数字转换为弧度	\
18	ROUND	将数字舍入为最接近的整数或指定小数位数	ROUND(3.1415,1)=3.1
19	SIGN	返回数字的符号：如果数字为正数，则返回1；如果数字为0，则返回0；如果数字为负数，则返回−1	\
20	SIN	返回角度的正弦，以弧度为单位指定角度	SIN(PI()/4)=0.707106781186548
21	SQRT	返回数字的平方根	SQRT(25)=5
22	SQUARE	返回给定数字的平方	AQUARE(5)=25
23	TAN	返回角度的正切，以弧度为单位指定角度	TAN(PI()/4)=1.0
24	ZN	如果<表达式>不为空，则返回它，否则返回零	ZN(Profit)

C.2　字符串函数

序　号	函　数	含　义	示　例
1	ASCII	返回字符串中第一个字符的ASCII代码值	ASCII("authors")=97
2	CHAR	将给定整数ASCII代码转换为字符	CHAR(65)='A'
3	CONTAINS	如果字符串包含子字符串，则返回True	CONTAINS("Calculation","alcu")为True
4	ENDSWITH	如果字符串以子字符串结尾(忽略尾随空格)，则返回True	ENDSWITH("Calculation","ion")为True
5	FIND	返回子字符串在字符串中的位置，如果未找到子字符串，则返回0。如果定义了起始参数，则会忽略在起始位置之前出现的所有子字符串实例。字符串中的第一个字符位置为1	FIND("Calculation","alcu")=2
6	ISDATE	如果给定字符串为有效日期，则返回True	1SDATE("2004-04-15")=True
7	LEFT	返回给定字符串开头的指定字符数	LEFT("Calculation",4)= "Calc"
8	LEN	返回给定字符串中的字符数	LEN("Calculation")=11
9	LOWER	将文本字符串转换为全小写字母	LOWER("ProductVersion")="productversion"
10	LTRIM	返回移除了所有前导空格的字符串	LTRIM(" Sales")= "Sales"
11	MAX	返回单个表达式在所有记录间的最大值，或两个表达式对于每个记录的最大值	MAX([Sales])

（续）

序　号	函　数	含　义	示　例
12	MID	在给定起始位置和长度的情况下，从文本字符串中间返回字符。字符串中的第一个字符位置为1。如果未包括长度，则将返回到字符串结束的所有字符。如果包括了长度，则最多返回该数量的字符	MID("Tableau Software",9)= "Software", MID("Tableau Software",2,4)= "able"
13	MIN	返回一个表达式在所有记录间的最小值，或两个表达式对于每个记录的最小值	MIN([Profit])
14	REPLACE	返回一个字符串，在该字符串中，子字符串的每次出现都会替换为替换字符串。如果未找到子字符串，则字符串保持不变	Replace("Calculation","ion","ed")=" Calculatied"
15	RIGHT	从给定字符串结尾起返回指定数量的字符	Right("Calculation",4)= "tion"
16	RTRIM	返回移除了所有尾随空格的字符串	RTRIM("Market ")="Market"
17	SPACE	返回由指定数量的重复空格组成的字符串	SPACE(2)= " "
18	STARTSWITH	如果字符串以子字符串开头，则返回True	STARTSWITH("Calculation","Ca")为True
19	TRIM	返回移除了前导和尾随空格的字符串	TRIM(" Budget ")="Budget"
20	UPPER	将文本字符串转换为全大写字母	UPPER("productversion")="PRODUCTVERSION"

C.3　日期函数

Tableau 提供多种日期函数。其中部分日期函数带有常量字符串参数date_part（参见表C-1），在使用日期函数时，请确保对应的date_part和表C-1一致。

C-1　日期函数表

date_part	值
'year'	四位数年份
'quarter'	1~4
'month'	1~12 或 "January"、"February" 等
'dayofyear'	一年中的第几天；1月1日为1、2月1日为32，依次类推
'day'	1~31
'weekday'	1~7 或 "Sunday"、"Monday" 等
'week'	1~52
'hour'	0~23
'minute'	0~59
'second'	0~60

1. DATEADD

函数公式为DATEADD(date_part, increment, date)，表示返回 date 增加increment后的日期，增加的程度由参数date_part决定。

例如，DATEADD('month',4,#6/24/2014#)返回结果为#10/24/2014#，即将日期6/24/2014加上4个月。如果为DATEADD('day',4,#6/24/2014#)则为加上4天，结果为#6/28/2014#。

2. DATEDIFF

函数公式为DATEDIFF(date_part, date1, date2, start_of_week)，表示返回 date1 与 date2 之差（以 date_part 的单位表示）。start_of_week 参数是可选参数，如果省略，一周的开始则由数据源确定。

例如，DATEDIFF('week', #2014-8-22#, #2014-08-24#, 'Sunday')= 0 ，#2014-8-22#为周三，#2014-08-24#为星期天，星期天为开始时，两者属于同一周，因此按照周来算差值为0。再如DATEDIFF('week', #2014-8-22#, #2014-08-24#, 'Friday')= 1，以星期五为开始，则星期三和星期天属于不同的周，因此为1。

3. DATENAME

函数公式为DATENAME(date_part, date, start_of_week)，表示以字符串的形式返回 date 的 date_part。start_of_week 参数是可选参数。

例如，DATENAME('month', #2014-08-15#) = "August"返回的为字符串August而不是8，注意与下面DATEPART的区别。

4. DATEPART

函数公式为DATEPART(date_part, date, start_of_week)，表示以整数形式返回 date 的 date_part。start_of_week 参数是可选参数。如果忽略start_of_week，则周起始日由为数据源配置的起始日决定。当 date_part 为weekday时会忽略 start_of_week 参数，这是因为 Tableau 依赖固定周日期顺序来应用偏移。

例如DATEPART('month', #2014-08-15#) =8。

5. DATEPARSE

函数公式为DATEPARSE(format, string)，将字符串转换为指定格式的日期时间。该函数并不常用，并非适用于所有数据源。数据中出现的不需要解析的字母应该用一对单引号' '引起来。对于值之间没有分隔符的格式（如*MMddyy*），请验证它们是否按按期方式解析。该格式必须是常量字符串，而非字段值。如果数据与格式不匹配，则返回 Null。

例如 DATEPARSE ("h'h' m'm' s's'", "10h 5m 3s") = #10:05:03# 对于h、m、s为非解析字段，因此用一对单引号''包起来。

6. DATETRUNC

函数公式为DATETRUNC(date_part, date, start_of_week)，表示按 date_part 指定的准确

度截断指定日期，返回新日期。start_of_week 数是可选参数。如果省略，一周的开始由数据源确定。

例如，DATETRUNC('month', #2014-08-15#) = 2014-08-01 12:00:00 AM 以月份级别截断处于月份中间的日期时，此函数返回当月的第一天；以季节级别截断时，返回该月所处季节的首月首天，如DATETRUNC('quarter', #2014-08-15#) = 2014-07-01 12:00:00 AM。

7. DAY

函数公式为DAY(date)，表示以整数形式返回给定日期的日。

例如DAY(#2014-08-15#) = 15，返回2014-08-15所在的日为15。

8. MONTH

函数公式为MONTH(date)，表示以整数形式返回给定日期的月。

例如MONTH(#2014-08-15#) = 8，返回2014-08-15所在的月为8。

9. YEAR

函数公式为YEAR(date)，表示以整数形式返回给定日期的年。

例如YEAR (#2014-08-15#) = 2014，返回2014-08-15所在的年为2014。

10. ISDATE

函数公式为ISDATE(string)，表示如果给定字符串为有效日期，则返回 True。

例如ISDATE("August 15, 2014") = true。

11. MAKEDATETIME

函数公式为MAKEDATETIME(date, time)，表示将date 和 time进行合并，产生一个日期和时间的数据。该函数仅适合于MySQL连接。

例如MAKEDATETIME("2014-08-15", #07:59:00#) = #08/15/2014 7:59:00 AM# 。

12. MAX

函数公式为MAX(expression) 或 MAX(expr1, expr2)，通常应用于数字，不过也适用于日期；返回 expr1 和 expr2 中的较大值（expr1 和 expr2 必须为相同类型）。如果任一参数为 Null，则返回 Null。

例如MAX(#2004-01-01# ,#2004-03-01#) = 2004-03-01 12:00:00 AM 。

13. MIN

与MAX类似，只是返回较小值。

14. NOW

函数公式为NOW()，返回当前日期和时间。

例如NOW() = 2014-08-15 10:43:21 AM 。

15. TODAY

函数公式为TODAY()，返回当前日期。

例如TODAY() = 2014-08-15 。

C.4　类型转换

计算中任何表达式的结果都可以转换为特定数据类型。转换函数包括 STR()、DATE()、DATETIME()、INT() 和 FLOAT()。例如，如果要将浮点数（如 3.14）转换为整数，则可以编写INT(3.14)=3。也可以将布尔值转换为整数、浮点数或字符串，但不能将其转换为日期。True 为 1（整数）、1.0（浮点）或"1"（字符串），而 False 为0、0.0 或"0"。Unknown 映射到 Null。

序　号	函　数	含　义	示　例
1	DATE	函数公式为DATE(expression)，表示在给定数字、字符串或日期表达式的情况下返回日期	DATE("August 15, 2014") = #August 15, 2014#
2	DATETIME	函数公式为DATETIME(expression)，表示在给定数字、字符串或日期表达式的情况下返回日期时间	DATETIME(“August 15, 2014 07:59:00”) = August 15, 2014 07:59:00
3	FLOAT	函数公式为FLOAT(expression)，表示将expression转换为浮点数	FLOAT(3) = 3.000，FLOAT([当期值]) 将当期值字段中的每个值转换为浮点数
4	INT	函数公式为INT(expression)，表示将expression转换为整数，这里不是四舍五入，而是直接取整数部分。字符串转换为整数时会先转换为浮点数，然后再取整。	INT(8.0/3.0) = 2 INT(4.0/1.5) = 2 INT(0.50/1.0) = 0 INT(-9.7) = -9
5	STR	函数公式为STR(expression)，表示将expression转换为字符串	STR([当期值])会将度量当期值中的数字转换为字符串

C.5　逻辑函数

Tableau自定义字段中经常使用逻辑函数，旨在不同条件下返回不同的值。

1. IF

IF函数的常用表达形式为IF test THEN value END 、IF test THEN value1 ELSE value2 END 或 IF test1 THEN value1 ELSEIF test2 THEN value2 ELSEIF test3 THEN value3 ... END。

使用IF test THEN value END时，先对test进行判断，如果为True 则返回value，否则为Null。例如IF [当期值]>10000 then "好" END，该语句对源数据行记录依次进行判断，如果当期值大于10 000，则该行赋值为"好"，如果当期值小于或等于10 000，则该行值赋值为Null。

使用IF test THEN value1 ELSE value2 END时，先对test进行判断，如果为True则返回value1，否则返回value2。例如IF [当期值]>10000 THEN "好" ELSE "不好" END ，源数据某行数据当期值

为12 000，则赋值为"好"，若某行数据当期值为9000，则赋值为"不好"。

前面两个语句最终的结果只有两类，如果要表示结果为多类，则需要用IF test1 THEN value1 ELSEIF test2 THEN value2 ELSEIF test3 THEN value3 … END。例如，IF[当期值]>10000 THEN "好" ELSEIF [当期值]<=10000 AND [当期值]>5000 THEN "较好" ELSE "不好" END，表示当当期值大于10 000时，赋值为"好"，当期值为大于5000并且小于等于10 000时为"较好"，其余的为"不好"。注意这里"大于5000并且小于等于10000"的表达式为" [当期值]<=10000 AND [当期值]>5000"，而不是"5000<[当期值]<=10000"，因为Tableau在执行判断的时候先判断"5000<[当期值] "，这时返回值为布尔值True 或 False ，继续就变成True(或False)<=10000，出错。

有时候上述三种类型任然不能满足需求，比如在[当期值]>10000的条件下，如果同期值大于10 000则为"好1"，其他为"好2"，则可通过内嵌条件语句来实现。具体语句可表示为IF[当期值]>10000 THEN (IF [同期值]>10000 THEN "好1" ELSE "好2" END) ELSEIF [当期值]<=10000 AND [当期值]>5000 THEN "较好" ELSE "不好" END。

另外，test可以是字段与字段比较，字段与参数比较。比如 IF [当期值]>[同期值] THEN "增长" ELSE "不增长" END。

2. CASE

CASE函数的作用和IF一样，但更方便使用，尤其是在返回结果较多的情况下。其表达式格式为：

```
CASE expression
WHEN value1 THEN return1
WHEN value2 THEN return2
WHEN value3 THEN return3
ELSE return4
END
```

编写函数时当然不必换行，只是换行逻辑更清楚些。该表达式表示当expression= value1时返回return1、当expression= value2时返回return2，以此类推。它等价于IF expression= value1 THEN return1 ELSEIF expression= value2 THEN return2 ELSEIF expression= value3 THEN return3 ELSE return4 END。

例如：

```
CASE [省市]
WHEN "吉林" OR "辽宁" OR "黑龙江" THEN "东北"
WHEN "安徽" OR "江苏" OR "浙江" OR "福建" THEN "华东"
WHEN "四川" OR "江西" OR "河南" OR "湖北" OR "湖南" OR "重庆" THEN "华中"
END
```

该表达式将省市（部分）划分了不同的区域。

CASE 主要是判断某 expression 与某些值是否匹配，但如果条件中存在比较，则还是得用IF语句。比如当期值大于10 000为"好"，得用IF语句。

3. IIF

函数公式为IIF(test, value1, value2, [unknown])，其中test为逻辑判断表达式。test 必须

是布尔值：数据源中的布尔字段或使用运算符的逻辑表达式的结果（或 AND、OR 或 NOT 的逻辑比较）。如果 test 计算为 True，则 IIF 返回 value1 值，如果 test 计算为 False，则 IIF 返回 value2 值。布尔比较可能生成值既不是 True 也不是 False，通常是因为测试中存在 Null 值，这时IIF返回的最后一个参数[unknown]，如果省略此参数，则会返回 Null。

例如IIF（7>5，"7大于5"，"7不大于5")=#7大于5# 。再如IIF（[当期值]>[同期值]，"当期值大于同期值","当期值小于或等于同期值")，执行语句时，对每一条行记录对当期值和同期值进行比较：如果当期值大于同期值，即[当期值]>[同期值]为True，返回值为#当期值大于同期值#；若当期值小于或等于同期值，则[当期值]>[同期值]为False，返回值为#当期值小于或等于同期值#。

4. IFNULL

函数公式为IFNULL(expression1, expression2)，如果结果不为 Null，则 IFNULL 函数返回第一个表达式，否则返回第二个表达式。

例如IFNULL（[当期值],[同期值])=[当期值] 如果不为空，则返回当期值，如果为空，则返回[同期值]。这实际上等价于将当期值的空值填补为同期值。

5. ISDATE

函数公式为ISDATE(string)，返回布尔值，如果string可以转换为日期，则返回True，否则返回False。

例如ISDATE("1/1/2014") = TRUE ，ISDATE("1/2014") = FALSE。

6. ISNULL

函数公式为ISNULL(expression)，返回布尔值，如果表达式为空，则返回True，否则返回False。

7. Min

函数表达式MIN(expression) 或 MIN(expression1,expression2) ，表示返回一个表达式在所有记录间的最小值，或两个表达式对于每个记录的最小值。

C.6　聚合函数

拖放度量字段时，Tableau会对其进行聚合运算，默认的聚合运算为SUM，此外可以手动选择聚合方式如，平均值、标准差、中位数等。Tableau提供了聚合运算函数，用户可自定义聚合运算。

序　号	函　数	含　义
1	ATTR	函数公式为ATTR(expression)，如果对expression所有行都有一个值，则返回该值，否则返回星号，会忽略 Null 值
2	AVG	函数公式为AVG(expression)，返回expression中所有值的平均值。AVG 只能用于数字字段，计算时会忽略 Null 值

（续）

序　号	函　数	含　义
3	COUNT	函数公式为COUNT(expression)，返回expression中的项目数（重复的项目仍然要计数，当作不同处理）。不对 Null 值计数
4	COUNTD	函数公式为COUNTD(expression)，返回组中不同项目的数量（重复项目当作一个处理）。不对 Null 值计数
5	MAX	函数公式为MAX(expression)，返回expression在所有记录中的最大值。如果表达式为字符串值，则此函数返回按字母顺序定义的最后一个值
6	MIN	函数公式为MIN(expression)，返回expression在所有记录中的最小值。如果表达式为字符串值，则此函数返回按字母顺序定义的第一个值
7	MEDIAN	函数公式为MEDIAN(expression)，返回expression所有记录中的中位数。中位数只能用于数字字段。将忽略空值
8	PERCENTILE	函数公式为PERCENTILE(expression, number)，表示从给定表达式返回与指定数字对应的百分位值。数字必须为0~1（含0和1），例如0.66，并且必须是数字常量
9	STDEV	函数公式为STDEV(expression)，表示基于群体样本返回给定表达式中所有值的统计标准差
10	STDEVP	函数公式为STDEVP(expression)，表示基于有偏差群体返回给定表达式中所有值的统计标准差。
11	SUM	函数公式为SUM(expression)，返回表达式中所有值的总计。SUM 只能用于数字字段。会忽略 Null 值
12	VAR	函数公式为VAR(expression)，表示基于群体样本返回给定表达式中所有值的统计方差
13	VARP	函数公式为VARP(expression)，表示对整个群体返回给定表达式中所有值的统计方差

C.7　表计算函数

Tableau提供多种表计算函数，以方便我们自定义表计算。一般而言，表计算可以分为五种类型：RUNNING_X、WINDOW_X、RANK_X、SCRIPT_X以及其他。

1. TOTAL

函数公式为TOTAL(expression)，表示返回给定表达式（expression）的总计。例如，TOTAL(SUM([人工服务接听量]))，用于计算各自分区中全部行的SUM(人工服务接听量)。

2. SIZE

函数公式为SIZE()，表示返回分区中的行数。如果当前分区包含5行，则SIZE()=5。

3. PREVIOUS_VALUE

函数公式为PREVIOUS_VALUE(expression)，表示返回此计算在上一行中的值，如果当前行是

分区的第一行，则返回给定表达式。

例如，PREVIOUS_VALUE(SUM([人工服务接听量])，意味着计算人工服务接听量总计聚合在上一行中的值。

4. LOOKUP

函数公式为LOOKUP(expression, [offset])，表示返回目标行（指定为与当前行的相对偏移）中表达式的值。如果省略了offset，则可以在字段菜单上设置要比较的行，当无法确定目标行时，则此函数返回Null。

例如，在分区中计算LOOKUP(SUM([人工服务接听量]), 2)时，每行都会显示接下来两行的数据。这里我们说明一个常用的小技巧，可以使用FIRST() +和LAST() -作为相对于分区中第一行/最后一行的目标偏移量定义的一部分。比如，LOOKUP(SUM([人工服务接听量]),FIRST()+2)表示计算分区第三行中的聚合值。

5. LAST

函数公式为LAST()，表示返回从当前行到分区中最后一行的行数。

例如，在分区中计算LAST()时，最后一行与倒数第二行之间的偏移为1。

6. INDEX

函数公式为INDEX()，表示返回分区中当前行的索引，不包含与值有关的任何排序。

例如，在分区中计算INDEX()时，各行的索引分别为1、2、3、4等。

7. FIRST

函数公式为FIRST()，表示返回从当前行到分区中第一行的行数。

例如，分区中计算FIRST()时，该分区的第一行与第二行之间的偏移为−1。

8. RANK

函数公式为RANK(expression, ['asc' | 'desc'])，表示返回分区中当前行按照聚合表达式（expression）的竞争排名。该函数为相同的值分配相同的排名。可选的 'asc' 、 'desc' 参数用于指定升序或降序（默认为降序）。

例如，RANK(SUM([人工服务接听量]), 'asc')意味着依据“人工服务接听量”总计聚合进行排序，然后进行竞争排名。如，对于“人工服务接听量”总计聚合排序为(6, 9, 9, 14)，那么这组值的排名为(1, 2, 2, 4)。

注意：排名函数中会忽略Null，并且它们不进行编号，也不计入百分位排名计算的总记录数中。

9. RANK_DENSE

函数公式为RANK_DENSE(expression, ['asc' | 'desc'])，表示返回分区中当前行按照聚合表达式（expression）的密集排名。该函数为相同的值分配相同的排名，但不会向数字序列中插入间距。可选的 'asc'、'desc' 参数用于指定升序或降序（默认为降序）。

例如，RANK_DENSE(SUM([人工服务接听量]), 'asc')意味着依据“人工服务接听量”总计聚

合进行排序，然后进行密集排名。例如，对于“人工服务接听量”总计聚合排序为(6, 9, 9, 14)，那么这组值的排名为(1, 2, 2, 3)。

注意：排名函数中会忽略Null，并且它们不进行编号，也不计入百分位排名计算的总记录数中。

10. RANK_MODIFIED

函数公式为RANK_MODIFIED(expression, ['asc' | 'desc'])，表示返回分区中当前行按照聚合表达式（expression）的调整后竞争排名。该函数为相同的值分配相同的排名。可选的 'asc'、'desc' 参数用于指定升序或降序（默认为降序）。

例如，RANK_ MODIFIED(SUM([人工服务接听量]), 'asc')意味着依据“人工服务接听量”总计聚合进行排序，然后进行调整后竞争排名。例如，对于“人工服务接听量”总计聚合排序为(6, 9, 9, 14)，那么这组值的排名为(1, 3, 3, 4)。

注意：排名函数中会忽略Null，并且它们不进行编号，也不计入百分位排名计算的总记录数中。

11. RANK_PERCENTILE

函数公式为RANK_PERCENTILE(expression, ['asc' | 'desc'])，表示返回分区中当前行按照聚合表达式（expression）的百分位排名。该函数为相同的值分配相同的排名。可选的 'asc'、'desc' 参数用于指定升序或降序（默认为降序）。

例如，RANK_PERCENTILE(SUM([人工服务接听量]), 'asc')意味着依据“人工服务接听量”总计聚合进行排序，然后进行百分位排名。如，对于“人工服务接听量”总计聚合排序为(6, 9, 9, 14)，那么这组值的排名为(25, 75, 75, 100)。

注意：排名函数中会忽略Null，并且它们不进行编号，也不计入百分位排名计算的总记录数中。

12. RANK_UNIQUE

函数公式为RANK_UNIQUE(expression, ['asc' | 'desc'])，表示返回分区中当前行按照聚合表达式（expression）的唯一排名。该函数为相同的值分配相同的排名。可选的 'asc'、'desc' 参数用于指定升序或降序（默认为降序）。

例如，RANK_UNIQUE(SUM([人工服务接听量]), 'asc')意味着依据“人工服务接听量”总计聚合进行排序，然后进行唯一排名。例如，对于“人工服务接听量”总计聚合排序为(6, 9, 9, 14)，那么这组值的排名为(1, 2, 3, 4)。

注意：排名函数中会忽略Null，并且它们不进行编号，也不计入百分位排名计算的总记录数中。

13. RUNNING_AVG

函数公式为RUNNING_AVG(expression)，表示返回给定聚合表达式（expression）从分区中第一行到当前行的平均值。

例如，沿着1月1日到1月31日的日期计算RUNNING_AVG(SUM([人工服务接听量])，则1月4日的计算方式为先计算1月4日的SUM([人工服务接听量])总计聚合值，然后计算从1月1日到1月4日的总计聚合值得平均值。

注意：与WINDOW_AVG不同，该方法只能计算第一行到当前行的运行平均值。

14. RUNNING_COUNT

函数公式为RUNNING_COUNT(expression)，表示返回给定聚合表达式（expression）从分区中第一行到当前行的计数。

例如，沿着1月1日到1月31日的日期计算RUNNING_COUNT(SUM([人工服务接听量])，则计算从1月1日到1月4日的总计聚合值的计数，此处为4。

15. RUNNING_MAX

函数公式为RUNNING_MAX(expression)，表示返回给定聚合表达式（expression）从分区中第一行到当前行的最大值。

例如，沿着1月1日到1月31日的日期计算RUNNING_MAX(SUM([人工服务接听量])，则1月4日的计算方式为先计算1月4日的SUM([人工服务接听量])总计聚合值，然后计算从1月1日到1月4日的总计聚合值中的最大值。

16. RUNNING_MIN

函数公式为RUNNING_MIN(expression)，表示返回给定聚合表达式（expression）从分区中第一行到当前行的最小值。

例如，沿着1月1日到1月31日的日期计算RUNNING_MIN(SUM([人工服务接听量])，则1月4日的计算方式为先计算1月4日的SUM([人工服务接听量])总计聚合值，然后计算从1月1日到1月4日的总计聚合值中的最小值。

17. RUNNING_SUM

函数公式为RUNNING_SUM(expression)，表示返回给定聚合表达式（expression）从分区中第一行到当前行的总计，简单点说就是各个分区的累积求和。

例如，沿着1月1日到1月31日的日期计算RUNNING_SUM(SUM([人工服务接听量])，则1月4日的计算方式为先计算1月4日的SUM([人工服务接听量])总计聚合值，然后计算从1月1日到1月4日的总计聚合值进行求和。

18. SCRIPT_BOOL

函数公式为SCRIPT_BOOL("R Script with arguments", .arg1, .arg2, …)，表示返回指定R表达式的布尔结果。

注意：书写包含占位符的R语言脚本后，我们可以通过在R表达式中使用“.arg*n*”的方式引用来自Tableau的参数（.arg1、.arg2 等），并且来自Tableau的传入参数要求聚合后的Tableau字段。

例如SCRIPT_BOOL("is.finite(.arg1)", SUM([人工服务接听量]))，其中R语言中is.finite()用于检测数据是否为无穷，在这里我们检测SUM([人工服务接听量])是否为无穷，返回True或者False。

19. SCRIPT_INT

函数公式为SCRIPT_INT("R Script with arguments", .arg1, .arg2, …)，表示返回指定R表

达式的整数结果。

注意：书写包含占位符的R语言脚本后，我们可以通过在R表达式中使用“.argn”的方式引用来自Tableau的参数（.arg1、.arg2 等），并且来自Tableau的传入参数要求聚合后的Tableau字段。

例如SCRIPT_INT("is.finite(.arg1)", SUM([人工服务接听量]))，其中R语言中is.finite()用于检测数据是否为无穷，在这里我们检测SUM([人工服务接听量])是否为无穷，但是返回值为1或者0。

20. SCRIPT_REAL

函数公式为SCRIPT_REAL("R Script with arguments", .arg1, .arg2, …)，表示返回指定R表达式的实数结果。

注意：书写包含占位符的R语言脚本后，我们可以通过在R表达式中使用“.arg*n*”的方式引用来自Tableau的参数（.arg1、.arg2 等），并且来自Tableau的传入参数要求聚合后的Tableau字段。

例如SCRIPT_REAL("is.finite(.arg1)", SUM([人工服务接听量]))，其中R语言中is.finite()用于检测数据是否为无穷，在这里我们检测SUM([Profit])是否为无穷，但是返回值为1.0或0.0。

21. SCRIPT_STR

函数公式为SCRIPT_STR("R Script with arguments", .arg1, .arg2, …)，表示返回指定R表达式的整数结果。

注意：书写包含占位符的R语言脚本后，我们可以通过在R表达式中使用“.arg*n*”的方式引用来自Tableau的参数（.arg1、.arg2等），并且来自Tableau的传入参数要求聚合后的Tableau字段。

例如SCRIPT_STR("is.finite(.arg1)", SUM([人工服务接听量]))，其中R语言中is.finite()用于检测数据是否为无穷，在这里我们检测SUM([人工服务接听量])是否为无穷，但是返回值为t或f。

22. WINDOW_AVG

函数公式为WINDOW_AVG (expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的平均值。

注意：我们可以使用 FIRST()+n 和 LAST()-n 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_AVG(SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的平均值。

23. WINDOW_COUNT

函数公式为WINDOW_ COUNT (expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的计数。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_ COUNT (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的计数。

24. WINDOW_MEDIAN

函数公式为WINDOW_ MEDIAN (expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的中位数。

注意：我们可以使用 FIRST()+和 LAST()-表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_ MEDIAN (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的中位数。

25. WINDOW_MAX

函数公式为WINDOW_ MAX (expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的中位数。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如，沿着分区向下WINDOW_ MAX (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的最大值。

26. WINDOW_MIN

函数公式为WINDOW_ MIN (expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的中位数。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_ MAX (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的最小值。

27. WINDOW_PERCENTILE

函数公式为WINDOW_PERCENTILE(expression, number, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）下指定百分位（number）相对应的值。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_PERCENTILE(SUM([人工服务接听量]), 0.75, -2, 0))，表示计算各个分区中当前行的前两行到当前行范围内的第75个百分位数。

28. WINDOW_STDEV

函数公式为WINDOW_ STDEV (expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的样本标准差。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_ STDEV (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的样本标准差。

29. WINDOW_STDEVP

函数公式为WINDOW_ STDEVP (expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的有偏差标准差。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_ STDEVP (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的有偏差标准差。

30. WINDOW_SUM

函数公式为WINDOW_ SUM (expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的总计求和。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_ SUM (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的总计求和。

31. WINDOW_VAR

函数公式为WINDOW_ VAR(expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的样本方差。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_ VAR (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的样本方差。

32. WINDOW_VARP

函数公式为WINDOW_ VARP(expression, [start, end])，表示返回窗口中从给定开头（start）到给定结尾（end）范围内给定表达式（expression）的有偏差方差。

注意：我们可以使用 FIRST()+ 和 LAST()- 表示与分区中第一行或最后一行的偏移，并且如果省略了开头（start）和结尾（end），则使用整个分区。

例如沿着分区向下WINDOW_ VARP (SUM[人工服务接听量]), FIRST()+1, 0)，表示计算各个分区中从第二行到当前行的SUM(人工服务接听量)的有偏差方差。

C.8 详细级别表达式

1. FIXED

表达式公式为{FIXED[dim1[,dim2]…]：aggregate-expression}，表示该表达式依据[dim1[,dim2]…]等维度进行聚合操作，而不论可视化内容中所用的其他维度如何。

例如，{Fixed[商品ID],[年份]:SUM([销量])}计算的是每件商品每年的销量总和，{Fixed[商品ID]:SUM([销量])}计算的是每件商品所有年份的销量总和。

2. INCLUDE

表达式公式为{INCLUDE[dim1[,dim2]…]：aggregate-expression}，使用指定的维度[dim1[,dim2]…]和视图维度进行计算聚合。INCLUDE关键字可创建聚合度低于（即粒度较高）可视化详细级别的表达式。

例如，{INCLUDE[销售人员]:SUM[销量]}计算每个销售人员处理订单的销量总和。[销售人员]这一维度在原来的可视化视图中可能并不存在，INCLUDE在计算聚合时将其纳入。

3. EXCLUDE

表达式公式为{EXCLUDE[dim1[,dim2]…]：aggregate-expression}，即如果指定的维度[dim1[,dim2]…]出现在视图中，则在计算聚合时会排除这些维度。EXCLUDE关键字可创建聚合度高于（即粒度较低）可视化详细级别的表达式。

例如，{EXCLUDE[地区]:SUM[销量]}计算的是根据已有的可视化视图的维度（例如有[商品],[年份]，[地区]），去除[地区]维度，计算销量总和，即计算每件商品每年的销量总和。

C.9 直通函数

RAWSQL 直通函数可用于将SQL表达式直接发送到数据库，而不由Tableau进行解析。如果你有 Tableau 不能识别的自定义数据库函数，则可以使用直通函数调用这些自定义函数。

你的数据库通常不会理解在 Tableau 中显示的字段名称。因为 Tableau 不会解释包含在直通函数中的 SQL 表达式，所以在表达式中使用 Tableau 字段名称可能会导致错误。可以使用替换语法将用于 Tableau 计算的正确字段名称或表达式插入直通 SQL。例如，假设有一个计算一组中值的函数，则可以对 Tableau 列 [Sales] 调用该函数，如下所示：

```
RAWSQLAGG_REAL("MEDIAN(%1)", [Sales])
```

因为 Tableau 不解释该表达式，所以必须定义聚合。在使用聚合表达式时，你可以使用下面描述的 RAWSQLAGG 函数。

RAWSQL 直通函数不使用已发布数据源。

在 Tableau Desktop 8.2 中启动时，这些函数可能返回与在较低版本的 Tableau Desktop 中不同的结果。这是因为 Tableau 现在对直通函数使用 ODBC 而非 OLE DB。当以整数形式返回实际值时，ODBC 会截断，而 OLE DB 会舍入。

Tableau 中提供了以下 RAWSQL 函数。

序　号	函　数	含　义	示　例
1	RAWSQL_BOOL	函数表达式 RAWSQL_BOOL("sql_expr", [arg1],...[a rgN])表示从给定的 SQL 表达式返回布尔结果。SQL 表达式直接传递给基础数据库。在 SQL 表达式中 % 用作数据库值的替换语法	表达%1 等于 [Sales]，%2 等于 [Profit]：RAWSQL_BOOL("IIF(%1→%2, True, False)", [Sales], [Profit])
2	RAWSQL_DATE	函数表达式 RAWSQL_DATE("sql_expr", [arg1],...[ar gN])表示从给定的 SQL 表达式返回日期结果。SQL 表达式直接传递给基础数据库。在 SQL 表达式中 %n 用作数据库值的替换语法	表达 %1 等于 [Order Date]：RAWSQL_DATE ("%1", [Order Date])
3	RAWSQL_DATETIME	函数表达式 RAWSQL_DATETIME("sql_expr", [arg1],...[argN])表示从给定的 SQL表达式返回日期和时间结果。SQL 表达式直接传递给基础数据库。SQL 表达式中 %n 用作数据库值的替换语法	表达 %1 等于 [Delivery Date]：RAWSQL_DATETIME("MIN(%1) ",[Delivery Date])
4	RAWSQL_INT	函数表达式 RAWSQL_INT ("sql_expr", [arg1],...[argN])表示从给定的 SQL 表达式返回整数结果。SQL 表达式直接传递给基础数据库。在 SQL 表达式中将 %n 用作数据库值的替换语法	表达%1 等于 [Sales]：RAWSQL_INT ("500 + %1", [Sales])
5	RAWSQL_REAL	函数表达式 RAWSQL_REAL("sql_expr", [arg1],...[argN])表示从直接传递给基础数据库的给定 SQL 表达式返回数字结果。在 SQL 表达式中 %n 用作数据库值的替换语法	例如表达%1 等于 [Sales]：RAWSQL_REAL("-123.98 * %1", [Sales])
6	RAWSQL_STR	函数表达式 RAWSQL_STR("sql_expr", [arg1],...[argN])表示从直接传递给基础数据库的给定 SQL 表达式返回字符串。在 SQL 表达式中 %n 用作数据库值的替换语法	表达 %1 等于 [Customer Name]：RAWSQL_STR("%1", [Customer Name])
7	RAWSQLAGG_BOOL	函数表达式 RAWSQLAGG_BOOL("sql_expr", [arg1],...[argN])表示从给定聚合 SQL 表达式返回布尔结果。SQL 表达式直接传递给基础数据库。在 SQL 表达式中将 %n 用作数据库值的替换语法	表达%1 等于 [Sales]，%2 等于 [Profit]：RAWSQLAGG_BOOL("SUM(%1) >SUM(%2)", [Sales], [Profit])
8	RAWSQLAGG_DATE	函数表达式 RAWSQLAGG_DATE("sql_expr", [arg1],...[argN])表示从给定的聚合 SQL 表达式返回日期结果。SQL 表达式直接传递给基础数据库。在 SQL 表达式中%n 用作数据库值的替换语法	比如表达%1 等于 [Order Date]：RAWSQLAGG_DATE(“MAX(%1)” ,[Order Date])
9	RAWSQLAGG_DATETIME	函数表达式 RAWSQLAGG_DATETIME("sql_ expr", [arg1],...[argN])表示从给定的聚合 SQL 表达式返回日期和时间结果。SQL 表达式直接传递给基础数据库。在 SQL 表达式中 %n 用作数据库值的替换语法	比如表达%1 等于 [Delivery Date]：RAWSQLAGG_DATETIME(“ MIN(%1) ” , [Delivery Date])

（续）

序　号	函　数	含　义	示　例
10	RAWSQLAGG_INT	函数表达式RAWSQLAGG_INT("sql_expr", arg1,...argN)表示从给定聚合 SQL 表达式返回整数结果。SQL 表达式直接传递给基础数据库。在 SQL 表达式中 %n 用作数据库值的替换语法	比如表达%1 等于 [Sales]：RAWSQLAGG_ INT("500 + SUM(%1)", [Sales])
11	RAWSQLAGG_REAL	函数表达式RAWSQLAGG_REAL("sql_expr", arg1,...argN)表示从直接传递给基础数据库的给定聚合 SQL 表达式返回数字结果。在 SQL 表达式中 %n 用作数据库值的替换语法	比如表达%1 等于 [Sales]：RAWSQLAGG_ REAL("SUM(%1) ", [Sales])
12	RAWSQLAGG_STR	函数表达式RAWSQLAGG_STR("sql_expr", arg1,...argN)表示从直接传递给基础数据库的给定聚合 SQL 表达式返回字符串。在 SQL 表达式中 %n 用作数据库值的替换语法	比如表达%1 等于 [Customer Name]：RAWSQLAGG_STR("AVG(%1)", [Discount])

C.10　用户函数

使用这些用户函数创建基于数据源中用户列表的用户筛选器。例如，假设你创建了一个视图，该视图显示有每个员工的销售业绩。发布该视图时，你可能希望仅允许员工查看自己的销售数据。这时可以使用函数CURRENTUSER 创建一个字段，该字段会在登录到服务器的人员的用户名与视图中的员工姓名相同时返回 True。然后，在使用此计算字段筛选视图时，视图只会显示当前已登录用户的数据。

1. FULLNAME

函数公式为FULLNAME()，返回当前用户的全名。当用户已登录时，这是 Tableau Server 或 Tableau Online 全名；否则为 Tableau Desktop 用户的本地或网络全名。比如[Manager]=FULLNAME()。

如果经理 Dave Hallsten 已登录，则仅当视图中的"Manager"字段包含"Dave Hallsten"时，此示例才会返回 True。用作筛选器时，此计算字段可用于创建用户筛选器，该筛选器仅显示与登录到服务器的人员相关的数据。

2. ISFULLNAME

函数表达式为ISFULLNAME(string)，表示如果当前用户的全名与指定的全名匹配，则返回 True；如果不匹配，则返回 False。当用户已登录时，此函数使用 Tableau Server 或 Online 全名；否则使用 Tableau Desktop 用户的本地或网络全名。

例如ISFULLNAME(“Dave Hallsten”)，如果 Dave Hallsten 为当前用户，则此示例返回 True，否则返回 False。

3. ISMEMBEROF

函数公式为ISMEMBEROF(string)，表示如果当前使用 Tableau 的人员是与给定字符串匹配的

组的成员，则返回 True。如果当前使用 Tableau 的人员已登录，则组成员身份由 Tableau Server 或 Tableau Online 上的组确定。如果该人员未登录，此函数返回 False。

例如IF ISMEMBEROF(“Sales”) THEN “Sales” ELSE “Other” END。

4. ISUSERNAME

函数表达式为ISUSERNAME(string)，如果当前用户的用户名与指定的用户名匹配，则返回 True；如果不匹配，则返回 False。当用户已登录时，此函数使用 Tableau Server 或 Online 用户名；否则使用 Tableau Desktop 用户的本地或网络用户名。

比如ISUSERNAME("dhallsten")，如果 dhallsten 为当前用户，则此示例返回 True，否则返回 False。

5. USERDOMAIN

函数表达式为USERDOMAIN()，当前用户已登录到 Tableau Server 时返回该用户的域。如果 Tableau Desktop 用户在域上，则返回 Windows 域；否则，此函数返回一个空字符串。

例如[Manager]=USERNAME() AND [Domain]=USERDOMAIN() 。

6. USERNAME

函数公式为USERNAME()，返回当前用户的用户名。当用户已登录时，这是 Tableau Server 或 Tableau Online 用户名；否则为 Tableau Desktop 用户的本地或网络用户名。

比如[Manager]=USERNAME()，如果经理 dhallsten 已登录，则仅当视图中的"Manager"字段为"dhallsten"时，此函数才会返回 True。用作筛选器时，此计算字段可用于创建用户筛选器，该筛选器仅显示与登录到服务器的人员相关的数据。

附录 D

数据表

本书所使用的数据表主要有“2014年各省市售电量明细表”（第2章以及4.1节、4.3节、4.4节、4.8节、4.9节、4.10节、5.1节）、“公司年龄统计表”（4.2节）、“2014年上半年综合计划指标明细表”（4.5节、4.6节、4.7节）、“物资采购情况明细表”（4.11节）、“蒙东地理数据”（5.3.1节）、“分供电所线损率数据”（5.3.2节）、“变电站数据”（5.3.3节）、“坐席接听”（第6章、第7章的7.2节、7.4节，以及第8章、第10章）、“物资采购金额”（7.1节）、“影响折旧费”（7.3节）、“全员劳动生产率指标”（7.5节）、“变压器、输电线路负载情况”（7.6节）、“坐席接听统计数据”（6.2.1节、6.3.1节、6.6.1节、7.2.1节、7.4节、8.1.1节、8.3.1节、8.4.3节）。数据源可以从图灵社区本书主页http://www.ituring.com.cn/book/1736免费下载使用。

D.1　2014 年各省市售电量明细表

该表来自于管理驾驶舱，指标为售电量，统计周期为2014年01月~2014年06月，数据存储为Excel文件。

数据表结构如图D-1所示，总共有9列变量，分别为省市、地市、统计周期、用电类别、当期值、累计值、同期值、同期累计值、和月度计划值。其中，用电类别是对售电量市场的进一步细分：包括大工业、居民、非居民、商业等9类，当期值为当月实际售电量，累计值为累计售电量，同期值为上一年相同月份的售电量，同期累计值为累计同期售电量，月度计划值为当月的计划值。

	A	B	C	D	E	F	G	H	I
1	省市	地市	统计周期	用电类别	当期值	累计值	同期值	同期累计值	月度计划值
2	重庆	重庆市区	1/31/2014	大工业	38567.77	38567.77	37153.4006	37153.4006	38567.77
3	重庆	江北	1/31/2014	大工业	24650.62	24650.62	22143.344	22143.344	24857.33

图D-1　2014年各省市售电量明细表

D.2　公司年龄统计表

数据存储为Excel文件。

数据表结构如图D-2所示，总共有3列变量，分别为出生日期、员工编号和年龄。

	A	B	C
1	员工编号	出生日期	年龄
2	80035756	1969/5/23	45
3	80035758	1968/4/12	46

图D-2　公司年龄统计表

D.3　2014 年上半年综合计划指标明细表

统计周期为2014年01月~2014年06月。数据存储为Excel文件。

数据表结构如图D-3所示。总共有4列变量，分别为指标名称、省市、统计周期和当期值。指标名称的值为售电量或利润总额，当期值为当月实际售电量或利润总额。

	A	B	C	D
1	指标名称	省市	统计周期	当期值
2	售电量	天津	1/31/2014	494262.5484
3	售电量	北京	2/28/2014	38042.05054

图D-3　2014年上半年综合计划指标明细表

D.4　物资采购情况明细表

数据存储为Excel文件。

数据表结构如图D-4所示，总共有5列变量，分别为采购订单号、物资类别、供应商名称、计划交货日期和实际交货日期。其中，物资类别包括交流变压器、交流断路器和交流隔离开关3类。

	A	B	C	D	E
1	采购订单号	物资类别	供应商名称	计划交货日期	实际交货日期
2	00000247	交流变压器	北京ABB高压开关设备有限公司	2014/1/20	2014/1/25
3	40308856	交流变压器	北京ABB高压开关设备有限公司	2014/3/15	2014/3/15

图D-4　物资采购情况明细表

D.5　蒙东地理数据

数据存储为Excel文件。

数据表结构如图D-5所示，总共有5列变量，分别为省市、地市、顺序、Latitude和Longitude。其中，省市的值为蒙东，地市是对蒙东的进一步划分：包括呼伦贝尔、兴安盟、赤峰市和通辽市4个地市，顺序为Tableau连接区域边界各点的顺序，Latitude为区域边界各点的纬度，Longitude为区域边界各点的经度。

	A	B	C	D	E
1	省市	地市	顺序	Latitude	Longitude
2	蒙东	通辽市	1	45.2911	119.2346
3	蒙东	通辽市	2	45.4724	119.3225

图D-5　蒙东填充地图数据

D.6　分供电所线损率数据

数据存储为Excel文件。

数据表结构如图D-6所示，总共有供电所、线损率两列数据。

	A	B	C	D
1	供电所	线损率	Latitude	Longitude
2	长港	0.031009694	343.8	940.73
3	杜山	4.348327652	608.55	1101.69

图D-6　分供电所线损率数据

D.7　变电站数据

数据存储为Excel文件。

数据表结构如图D-7所示，总共有6列数据，分别为国家/地区、省/市/自治区、城市、变电站、Longitude和Latitude。

	A	B	C	D	E	F
1	国家/地区	省/市/自治区	城市	变电站	Longitude	Latitude
2	中国	重庆	南岸	变电站A	106.56	29.52
3	中国	重庆	南岸	变电站B	106.57	29.54

图D-7　变电站数据

D.8　物资采购金额

数据存储为Excel文件。

数据表结构如图D-8所示，总共有2列数据，分别为供应商名称、应付金额。

	A	B
1	供应商名称	应付金额
2	1	52.5447
3	2	33.67728
4	3	76.3323795

图D-8　物资采购金额

D.9　影响折旧费

数据存储为Excel文件。

数据表结构如图D-9所示，总共有3列数据，分别为单位名称、单位类别和影响折旧费。其中，单位类别按各条记录影响折旧费的正负分为正影响折旧费单位和负影响折旧费单位。

	A	B	C
1	单位名称	单位类别	影响折旧费（万元）
2	漳州	正影响折旧费单位	686.4689773
3	厦门	负影响折旧费单位	-650.7448077

图D-9　影响折旧费

D.10　全员劳动生产率指标

数据存储为Excel文件。

数据表结构如图D-10所示，总共有4列数据，分别为指标名称、单位、当期累计值和同期累计值。其中，指标名称的值为全民劳动生产率。

	A	B	C	D
1	指标名称	单位	当期累计值	同期累计值
2	全员劳动生产率	上海	33	41.83
3	全员劳动生产率	北京	26	44.92

图D-10　全员劳动生产率指标

D.11　变压器、输电线路负载情况

数据存储为Excel文件。

数据表结构如图D-11所示，总共有11列数据，分别为线路、变电站、*X*、*Y*、变电站所属单位、变电站数量、#1主变、#2主变和#3主变、负载率、投运年限。

	A	B	C	D	E	F	G	H	I	J	K
1	线路	变电站	X	Y	变电站所属单位	变电站数量	#1主变	#2主变	#3主变	负载率	投运年限
2	500kV云天线	t	325.26	128.83	清远供电局	0	0%	0%	0%	84%	8
3	500kV云天线	v	357.98	139.25	清远供电局	0	0%	0%	0%	84%	8

图D-11　变压器、输电线路负载

D.12　坐席接听统计数据

坐席接听统计数据，统计周期为2014年1月1日至2014年1月31日（参见图D-12）。数据存储为EXCEL文件。其中每条记录包括一个坐席一天接听电话情况的统计信息，以及该坐席的所属

中心、部门、组、班、工号、姓名等基本信息。数据存储为Excel文件。

数据表结构如图。总共有15列变量，分别为中心、部、组、班、日期、姓名、工号、人工服务接听量、三声响铃接听量、呼入通话时长（秒）、平均呼入通话时长（秒）、呼入案头总时长（秒）、平均案头时长（秒）、服务评价推送成功数和服务评价满意率。其中，每位员工都有唯一的工号。

中心	部	组	班	日期	姓名	工号	人工服务接听	三声铃响接听	呼入通话时长	平均呼入通话	呼入案头总时	平均呼入案头	服务评价推送	服务评价满意
北中心	客服一部	客服一组	常白班	2014年1月1日	张三	0786	46	46	4513.00	98.11	224.00	4.87	41	95.12%
北中心	客服一部	客服二组	四班	2014年1月1日	李四	0502	58	58	6163.00	106.26	396.00	6.83	50	94.00%

图D-12　坐席接听统计数据

中心、部、组、班是层级关系，某中心下设多个部，某部下设多个组，某组下设多个班。中心包括南中心、北中心，部包括客服一部、客服二部和客服三部，组包括客服一组、客服二组，班包括常白班、新人班、1班、运行1班等共32个班。人工服务接听量为一个坐席一天接听电话数量，三声响铃接听量为电话铃响三声内接听的电话数量，呼入通话时长（秒）为一个坐席一天接听电话总通话时长，平均呼入通话时长（秒）为以为坐席一天内每则通话平均时长，呼入案头总时长（秒）为一个坐席一天内完成一次呼叫后到接听下一通电话之间的总时长，平均案头时长（秒）为一个坐席一天内呼入案头总时长（秒）除以该天的人工服务接听量。



“本书通过真实的案例，阐述了一个完全不同于以往的数据分析方法论。它展示了领先企业如何让商业智能不再局限于少数技术人员，让多数人都掌握自助分析，读懂数据创造更大的价值。”

——JY Pook， Tableau亚太区副总裁

“Tableau是一座桥梁，是一座将业务运营和大数据分析紧密融合的生产力桥梁。本书是一本以业务为导向、以Tableau 为生产工具的最好的可视化学图书，让人倍感亲切，不忍释卷。”

——柳卿博士，上海贝尔股份有限公司

“非常幸运能在第一时间读到这本书，对于Tableau和数据爱好者来说，这绝对是一本值得推荐的书。”

——栾超，特斯拉（中国）工程师

“Tableau 提供了几乎所有数据源的访问接口，分析处理能力强大，图表精美，操作简单。本书分析视角独特且多样，非常值得一读。”

——庄晓天博士，亚马逊（中国）Research Scientist

图灵社区：iTuring.cn
热线：(010) 51095186 转 600

分类建议 计算机/数据可视化
计算机/统计分析软件

人民邮电出版社网址：www.ptpress.com.cn

ISBN 978-7-115-40686-6

ISBN 978-7-115-40686-6
定价：69.00元

看完了

如果您对本书内容有疑问，可发邮件至 contact@turingbook.com，会有编辑或作译者协助答疑。也可访问图灵社区，参与本书讨论。

如果是有关电子书的建议或问题，请联系专用客服邮箱:
ebook@turingbook.com。

在这可以找到我们:

微博 @图灵教育：好书、活动每日播报
微博 @图灵社区：电子书和好文章的消息
微博 @图灵新知：图灵教育的科普小组
微信 图灵访谈：ituring_interview，讲述码农精彩人生
微信 图灵教育：turingbooks